

*Digital Comprehensive Summaries of Uppsala Dissertations
from the Faculty of Pharmacy 397*

Development and Application of Computational Methods in Mass Spectrometry Imaging

PATRIK BJÄRTEROT



ACTA UNIVERSITATIS
UPSALIENSIS
2026



UPPSALA
UNIVERSITET

Dissertation presented at Uppsala University to be publicly examined in BMC A1:111, Husargatan 3, Uppsala, Friday, 6 March 2026 at 13:15 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: Assistant Professor Benjamin Balluff (Maastricht University).

Abstract

Bjärterot, P. 2026. Development and Application of Computational Methods in Mass Spectrometry Imaging. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Pharmacy* 397. 56 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-513-2718-1.

Mass spectrometry imaging (MSI) is an emerging technique for spatially resolving the molecular composition of biological samples. MSI frequently relies on matrix-assisted laser desorption/ionization (MALDI), in which a pulsed laser beam and chemical matrices are used to facilitate desorption/ionization of molecular species from the sample surface. MALDI matrices can be divided into two broad groups: conventional matrices that promote ionization by protonation/deprotonation or cationization, and derivatizing matrices that target specific chemical functionalities. Derivatizing matrices such as FMP-10 are charged molecules that react with specific chemical structures on target analytes to form covalent matrix-analyte conjugates, enhancing ionization and detectability but limiting chemical coverage. Derivatizing matrices may also create multiple derivatization products through serial reactions with single analytes, complicating annotation. This prompted development of Met-ID, a software tool for automatic annotation of MSI data with an emphasis on derivatization-based workflows. Met-ID incorporates matrix-specific chemistry to enumerate plausible derivative products and filter chemically implausible annotations. It includes a database of in-house acquired tandem mass spectrometry (MS2) spectra of FMP-10-derivatized chemical standards to support MS2 spectral matching. The use of ion mobility (IM) spectrometry in MSI enables collision cross section (CCS) values to be used for annotation. This motivated the development of CCSSim, an in-silico CCS prediction method implemented in Met-ID together with a mixture-model framework to increase annotation confidence by integrating m/z and CCS data. To improve spatial correlations between mass spectrometric and transcriptomic data, a method was developed to enable sequential MSI and spatially resolved transcriptomics (SRT) analysis of one tissue section rather than using consecutive sections. This spatial multimodal analysis can be performed on non-conductive Visium slides without appreciable degradation of MSI metabolite signal or SRT RNA signal. Finally, MALDI-MSI was evaluated as a sample-efficient approach for distinguishing de novo Parkinson's disease patients from controls using limited patient material and minimal sample preparation, reducing analytical time compared to more sample-intensive workflows. In conclusion, this thesis introduces new high-throughput computational methods for automated metabolite annotation in tissue sections, demonstrates the compatibility of MALDI-MSI with SRT, and highlights the versatility of MSI for analyzing sample-limited clinical biofluids.

Keywords: analytical chemistry, software, bioinformatics, mass spectrometry imaging, spatial omics

Patrik Bjärterot, Department of Pharmaceutical Biosciences, Box 591, Uppsala University, SE-75124 Uppsala, Sweden.

© Patrik Bjärterot 2026

ISSN 1651-6192

ISBN 978-91-513-2718-1

URN urn:nbn:se:uu:diva-575210 (<http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-575210>)

For all the people that made this possible

List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

- I. **Bjärterot, P.**; Nilsson, A.; Shariatgorji, R.; Vallianatou, T.; Kaya, I.; Svenningsson, P.; Käll, L.; Andrén, P.E.; Met-ID: An Open-Source Software for Comprehensive Annotation of Multiple On-Tissue Chemical Modifications in MALDI-MSI. *Analytical Chemistry* **2025**, *97* (16), 9033-9041. DOI: 10.1021/acs.analchem.5c00633
- II. **Bjärterot, P.**; Vallianatou, T.; Nilsson, A.; Shariatgorji, R.; Kaya, I.; Svenningsson, P.; Käll, L.; Andrén, P.E.; Met-ID 2.0: High Confidence Annotation with Collision Cross Section Predictions. *In manuscript*. **2026**
- III. Vicari, M.; Mirzazadeh, R.; Nilsson, A.; Shariatgorji, R.; **Bjärterot, P.**; Larsson, L.; Lee, H.; Nilsson, M.; Foyer, J.; Ekvall, M.; Czarnewski, P.; Zhang, X.; Svenningsson, P.; Käll, L.; Andrén, P.E.; Lundeborg, J. Spatial multimodal analysis of transcriptomes and metabolomes in tissues. *Nature Biotechnology* **2024**, *42* (7), 1046-1050. DOI: 10.1038/s41587-023-01937-y.
- IV. Vallianatou, T.; Nilsson, A.; **Bjärterot, P.**; Shariatgorji, R.; Slijkhuis, N.; Aerts, J. T.; Jansson, E. T.; Svenningsson, P.; Andrén, P. E. Rapid Metabolic Profiling of 1 μ L Crude Cerebrospinal Fluid by Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry Imaging Can Differentiate De Novo Parkinson's Disease. *Analytical Chemistry* **2023**, *95* (50), 18352-18360. DOI: 10.1021/acs.analchem.3c02900.

Reprints were made with permission from the respective publishers.

List of Additional Papers

Kaya, I.; Nilsson, A.; Luptáková, D.; He, Y.; Vallianatou, T.; Bjärterot, P.; Svenningsson, P.; Bezard, E.; André, P.E.; Spatial lipidomics reveals brain region-specific changes of sulfatides in an experimental MPTP Parkinson's disease primate model. *NPJ Parkinsons Dis.* **2023** 26;9(1):118. DOI: 10.1038/s41531-023-00558-1.

Kaya, I.; Vallianatou, T.; Nilsson, A.; Bjärterot, P.; Shariatgorji, R.; Svenningsson, P.; Bezard, E.; André, P.E.; Brain-region-specific lipid dysregulation in L-DOPA-induced dyskinesia in a primate model of Parkinson's disease. *NPJ Parkinsons Dis.* **2025** 23;11(1):258. DOI: 10.1038/s41531-025-01109-6.

Mirzazadeh, R.; Myers, M.; Nilsson, A.; Shariatgorji, R.; Troung, P.; Morlanes J.E.; Bjärterot, P.; Vicari, M.; Shakari, N.; Nilsson M.; André, P.; Lundberg, J.; Spatial Multimodal Analysis: A workflow for Profiling Resolved Metabolomics, Transcriptomics and Histology in a Single Tissue Section. *Nat. Protocols.* In press

Contents

Introduction.....	11
Software for scientific research.....	11
Mass Spectrometry Imaging.....	12
MSI ionization techniques.....	13
MALDI-MSI.....	13
Mass Analyzers.....	16
Fourier transform ion cyclotron resonance.....	16
Time-of-flight.....	16
Other mass analyzers.....	17
Tandem Mass Spectrometry.....	17
Ion Mobility Spectrometry.....	17
Data collection and analysis.....	18
MSI software.....	18
Annotation and identification.....	19
Metabolite Annotation and Identification.....	19
Collision Cross Section.....	21
Multimodality.....	22
Scientific Programming.....	22
Artificial Intelligence in Computational Mass Spectrometry.....	25
Aims.....	27
Methods.....	28
Human Samples.....	28
Animal Samples Experiments.....	28
MSI workflow.....	29
Spatial Multimodal Analysis Workflow.....	30
Software and Programming languages.....	31
Results and Discussion.....	33
The Met-ID Software (Paper I & II).....	33
Development of Met-ID.....	35
The Met-ID metabolite identification algorithm.....	35
CCS prediction.....	37
Mixture model for identification confidence.....	38
The Met-ID REST-API.....	39
Validation of Met-ID.....	39

Spatial Multimodal Analysis (Paper III)	40
Spatial metabolomics and transcriptomics on the same tissue	40
Low-volume CSF Metabolomics for De Novo Parkinson's Disease Classification (Paper IV)	42
Minimal sample preparation and MALDI-MSI acquisition	42
Statistical modeling and feature selection to discriminate de novo Parkinson's disease from controls	43
Conclusions	44
Populärvetenskaplig sammanfattning	45
Acknowledgements	47
References	49
Generative AI	56

Abbreviations

9AA	9-Aminoacridine
AI	Artificial intelligence
API	Application programming interface
CCS	Collision cross section
CE-MS	Capillary electrophoresis mass spectrometry
CD	Continuous delivery
CI	Continuous integration
CID	Collision induced dissociation
CLI	Command line interface
CSF	Cerebrospinal fluid
DESI	Desorption electrospray ionization
DHB	2,5-Dihydroxybenzoic acid
eV	Electronvolt
FAIR	Findability, accessibility, interoperability, reusability
FMP	2-Fluoro-1-methyl pyridinium
FPS	Frames per second
FTICR	Fourier transform ion cyclotron resonance
GB	Gigabyte
GIL	Global interpreter lock
GUI	Graphical user interface
H&E	Hematoxylin and eosin
IMS	Ion mobility spectrometry
IM-MSI	Ion mobility spectrometry – mass spectrometry imaging
ISS	In situ sequencing
ITO	Indium tin oxide
KB	Kilobyte
LLM	Large language model
MALDI	Matrix-assisted laser desorption/ionization
MAPE	Mean absolute percentage error
mRNA	Messenger ribonucleic acid
MS	Mass spectrometry
MS2	Tandem mass spectrometry
MSI	Mass spectrometry imaging

NEDC	N-(1-naphthyl)ethylenediamine dihydrochloride
PCA	Principal components analysis
PD	Parkinson's disease
RAM	Random access memory
REST	Representational state transfer
RNA	Ribonucleic acid
SIMS	Secondary ion mass spectrometry
SMA	Spatial multimodal analysis
SRT	Spatially resolved transcriptomics
ST	Spatial transcriptomics
TAHS	p-N,N,N-Trimethylammonioanilyl N-hydroxysuccinimidyl carbamate iodide
TIMS	Trapped ion mobility spectrometry
TOF	Time-of-flight

Introduction

Software plays essential roles at every stage of the research process in modern scientific research, from experimental design and data acquisition to analysis and interpretation. Computational advances made in recent decades have transformed research methodologies, enabling automation and high-throughput data processing, and providing assistance with scientific writing [1, 2]. The evolution of software is exemplified by the transition from the handwritten/hand-assembled software created by Margaret Hamilton for the Apollo 11 mission in 1969 [3], to AlphaFold [4, 5], the artificial intelligence system that revolutionized protein structure prediction and won its main authors the Nobel prize in 2024 [6]. This technological progress has been accompanied by exponential hardware improvements; the Apollo 11 computer operated with kilobytes (KB) of random-access memory (RAM) whereas modern smartphones and laptops commonly feature gigabytes (GB) - a million-fold increase.

Software for scientific research

The development of software for scientific (and more specifically, biological) research, has evolved significantly from early computational tools like BLAST (1990) [7] for amino acid and nucleotide sequence comparison to modern neural networks capable of complex data analysis in genomics [8, 9], proteomics [4, 5] and metabolomics [10]. These advancements have been facilitated by the emergence and widespread adoption of user-friendly programming languages, open-source software, and generative AI, all of which have increased the accessibility of programming.

The introduction of Python (1991) [11], R (1993) [12], and Julia (2012) [13], has lowered the threshold to learn programming and allowed researchers to develop custom analytical tools and process large datasets without extensive expertise in computer science, fostering a more data-driven approach to biological research. The basics of programming and operating a terminal are now commonly taught in undergraduate and graduate science education, which streamlines the scientific process by allowing non-bioinformaticians to make small changes and run pipelines without consulting a programmer.

Another major breakthrough in biological software development stems from the shift towards open-source software and movements such as Findability, Accessibility, Interoperability and Reusability (FAIR) [14], which have helped scientists to openly share their work and collaborate on cutting-edge technology in real time. This process has been facilitated by code-sharing platforms such as Github [15], a free-to-use site for hosting project code and software pipelines.

Most recently, the development and availability of large language models (LLMs) such as Microsoft's Copilot [16] or OpenAI's ChatGPT [1] has further increased the accessibility of software development. LLMs perform very well at simple, well-documented tasks, which suits programming novices. However, they can be less reliable when applied to poorly documented or novel tasks which still require deep programming expertise.

Mass Spectrometry Imaging

Mass Spectrometry Imaging (MSI) [17] is an emerging technique for spatially resolving the molecular composition of biological samples. It requires specialized procedures for sample preparation and desorption/ionization as well as spatially resolved mass spectrometric analysis, performed in a spatially resolved manner [18, 19]. There are different types of MSI instruments that each have distinct advantages and disadvantages, but share the fundamental purpose of measuring and mapping ion distributions across a sample. MSI experiments generate large and complex datasets [20, 21] that necessitate the use of sophisticated data analysis methods.

The subject of this thesis is software development for spatial multi-omics analysis, with an emphasis on spatial metabolomics. It presents the development of several computational tools designed to address key challenges in MSI data processing and integration, including (a) automated annotation of MSI data, particularly for derivatizing matrices, based on the structural characteristics of target metabolites (Papers I & II); (b) multimodal spatial analysis, enabling the combined use of MSI and spatially resolved transcriptomics (SRT) on the same tissue section to improve molecular mapping (Paper III); and (c) minimal sample preparation workflows and metabolite annotation, demonstrated by the analysis of one μL of cerebrospinal fluid (CSF) for biomarker discovery (Paper IV).

These advancements contribute to the development of more efficient, automated, and integrated computational frameworks for MSI-based spatial metabolomics.

MSI ionization techniques

The foundations of MSI were established in the 1960s [22, 23] and 1970s [24] with the development of ionization techniques for conventional mass spectrometry. These techniques were subsequently adapted for spatially resolved analysis of biological tissues, leading to the emergence of MSI. The field gained significant momentum in the late 1990s when Caprioli et al. demonstrated the feasibility of imaging peptides and proteins using matrix-assisted laser desorption/ionization (MALDI)-MSI [17], marking a pivotal advancement in the application of mass spectrometry to spatial biomolecular analysis. Since then, three ionization techniques have been established for imaging molecules in biological tissues; MALDI, desorption electrospray ionization (DESI), nano-DESI, and secondary ion mass spectrometry (SIMS).

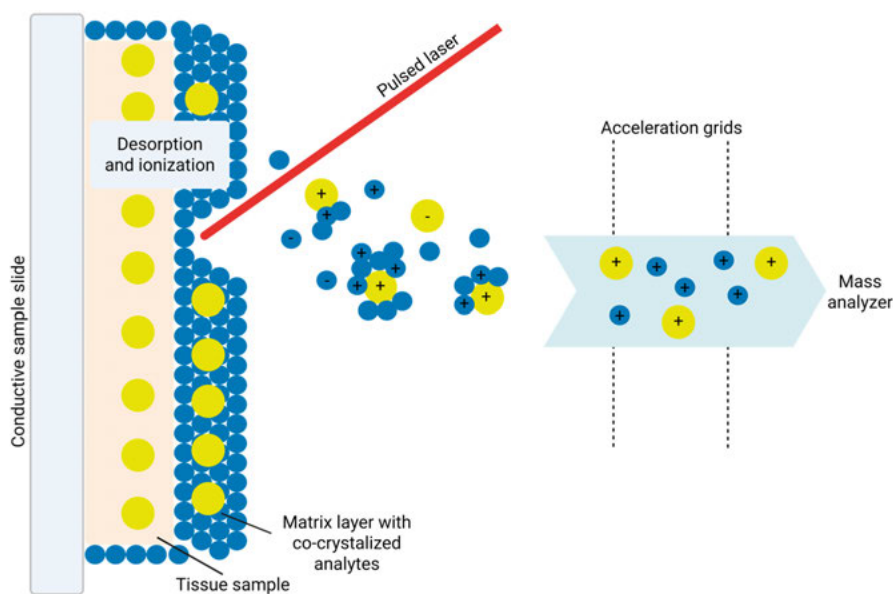


Figure 1. Schematic illustration of MALDI. Analytes are co-crystallized with an excess of MALDI matrix on the sample surface. Upon irradiation with a pulsed laser, the matrix strongly absorbs the laser energy and initiates rapid desorption of the matrix-analyte plume. Analyte molecules are ionized predominantly via proton transfer and cationization reactions, generating ions such as $[M+H]^+$, $[M+Na]^+$, or $[M+K]^+$ (or, in negative-ion mode, $[M-H]^-$ and related adduct ions). The resulting ions are extracted from the source by an applied electric field and accelerated through extraction/acceleration grids into the mass analyzer for m/z measurement.

MALDI-MSI

MALDI is the most widely used ionization technique in MSI [18]. MALDI-MSI uses a pulsed laser beam and chemical matrices to aid the desorption/ionization

of molecular species (Figure 1). MALDI matrices [25] can be divided into two categories based on their chemical behavior; conventional matrices are unspecific in their ionization [26], whereas derivatizing matrices are developed to target specific chemical structures [27-29].

Conventional matrices

Conventional matrices such as DHB [30, 31], CHCA [32, 33], 9AA [32, 34] and NEDC [27, 35] are used in either positive- or negative- ion modes and typically do not chemically derivatize the analyte; instead, they facilitate desorption/ionization and the formation of common ion species including $[M+H]^+$, $[M+K]^+$ and $[M+Na]^+$ in positive-ion mode or $[M-H]^-$ (and, depending on conditions, adducts such as $[M+K-2H]^-$, $[M+Na-2H]^-$ and $[M+Cl]^-$) in negative ion-mode.

Acidic functional groups such as fatty acids, nucleotides and sulfates are favored in negative ion mode, while alkaline functional groups such as small peptides and some lipids are favored in positive ion mode. Norharmane is an exception to this rule [36, 37]: it can be used in both negative and positive ion mode, favoring lipids in both.

Derivatizing matrices

Certain metabolites exhibit poor ionization efficiency with conventional matrices, which has led to the development of derivatizing agents/matrices including FMP-10 [38], TAHS [39], Girards T and P [40], and many more as reviewed by Calvano et al. [26] and Zhou et al. [41]. These matrices are charged molecules that react with target analytes to form larger matrix-analyte conjugates, enhancing ionization. Some derivatizing matrices, such as FMP-10 [38], incorporate two functional domains: a highly reactive moiety that enables covalent charge-tagging of the analyte, and a conjugated chromophore that facilitates laser desorption.

Target molecules may have multiple derivatization sites and may thus react with the matrix multiple times. Moreover, the derivatized products may undergo downstream reactions. Both factors complicate automated annotation. This complexity is exemplified by the derivatizing matrix FMP-10 [38], which is commonly used in neurotransmitter analysis and reacts with both phenolic hydroxyl groups and primary amine groups. Dopamine contains three such groups, so its reaction with FMP-10 produces several distinct peaks corresponding to single, double, and triple derivatization products (Figure 2). Furthermore, each derivatization reaction beyond the first may be followed by the loss of a methyl group.

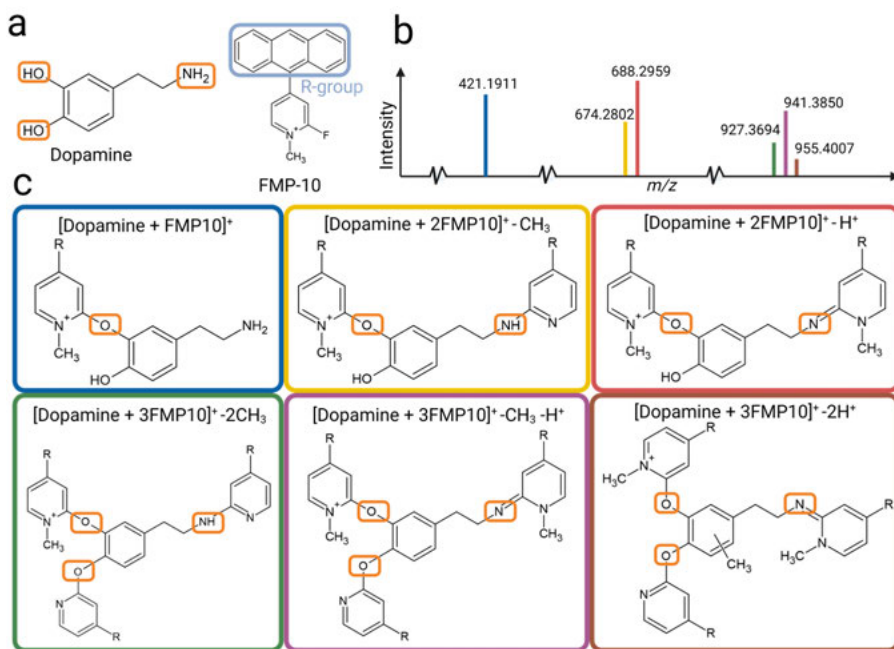


Figure 2. FMP-10 derivatization of dopamine. FMP-10 reacts with primary amines and phenolic hydroxyls, indicated with an orange box. Dopamine contains three derivatization sites corresponding to 6 different peaks factoring in FMP-10 derivatization chemistry. (a) Shows the chemical structures of dopamine and FMP-10, (b) shows the mass spectrum of FMP-10 derivatized dopamine and (c) shows example structures from each of the 6 peaks. The colors of the boxes correspond to the color of the peaks in the mass spectrum.

Other MSI ionization techniques

Prominent ionization techniques complementary to MALDI are DESI [42], nano-DESI [43], and SIMS [22]. These techniques do not require an added matrix, which simplifies experiments and reduces sample preprocessing. DESI uses an electrospray to desorb and ionize analytes from the sample surface, while nano-DESI uses a solvent microjunction (typically formed between two capillaries) to locally extract analytes from the surface and ionize them in an ESI-like manner. In contrast, SIMS uses a focused beam of primary ions such as metal ions [44] to sputter material from the surface and generate secondary ions from the tissue. DESI and nano-DESI are gentle techniques while SIMS is more destructive [45], but offers greater lateral resolution [45].

Mass Analyzers

Instruments for measuring the mass of different atomic and molecular species have existed for over a century. The foundations of modern mass spectrometry were laid in 1912 at the University of Cambridge, when J.J. Thomson used a positive-ray method with electric and magnetic fields to separate and measure isotopes, identifying two stable neon isotopes (^{20}Ne and ^{22}Ne) [46]. Building on his work, in 1919, his student F.W. Aston described and developed the first mass spectrometer, a positive ray spectrograph, enabling high-precision isotopic measurements [47]. A wide range of mass spectrometry instruments were developed over the following century, including Fourier transform ion cyclotron resonance (FT-ICR) and time-of-flight (TOF) instruments, each offering distinct performance advantages and methodological limitations.

Fourier transform ion cyclotron resonance

The FT-ICR mass spectrometer is a highly specialized instrument for high mass accuracy and resolution mass spectrometry [48, 49]. It functions by trapping ions in a magnetic field, where they undergo cyclotron motion. The resulting induced voltage over time, known as the ICR transient, is recorded and subsequently processed using a Fourier transform to convert cyclotron frequencies into a mass spectrum. The mass accuracy of FT-ICR-MS is directly influenced by the strength of the applied magnetic field, measured in Tesla; higher magnetic flux densities enable improved resolution and precision.

Time-of-flight

TOF [50] mass spectrometry is a widely used analytical technique that measures the time required for ions to travel through a flight tube to a detector. The ions' velocity is inversely proportional to the square root of their m/z ratio, meaning that lighter ions travel faster than heavier ones under the same acceleration conditions. TOF analyzers are valued for their high acquisition speed, which makes them particularly suitable for MSI applications that require rapid data collection across large tissue sections. Modern TOF instruments often incorporate reflectrons [51], i.e., electrostatic ion mirrors that improve mass resolution by correcting for small energy variations among ions of the same m/z , leading to sharper peak shapes and improved mass accuracy. Despite their advantages in speed and mass range, TOF-MS instruments generally have lower mass accuracy and resolving power than high-resolution instruments such as FT-ICR or Orbitrap analyzers, which can limit their ability to distinguish isobaric species.

Other mass analyzers

Other types of mass analyzers, not employed in this project, include ion trap-based instruments such as the widely used Orbitrap [52, 53]. Whereas FT-ICR mass spectrometry relies on a strong magnetic field, the Orbitrap uses electrostatic trapping in which ions oscillate harmonically around a central electrode. The oscillation frequency is inversely proportional to the ion's m/z ratio, with lighter ions having higher frequencies than heavier ones. While the Orbitrap does not require a superconducting magnet, making it more accessible and compact, this comes at the cost of worse mass resolution than FT-ICR-MS.

Tandem Mass Spectrometry

Tandem mass spectrometry (MS2) involves identifying ions by fragmentation. After analyzing the most interesting peaks in the MS1 spectra, ions within a selected isolation window are focused into the collision cell to be fragmented by collision-induced dissociation (CID). The isolation window defines the range of m/z ratios selected for fragmentation. Narrower isolation windows lead to higher specificity at the cost of lower experimental sensitivity and vice versa. The collision energy used in CID [54] is measured in electronvolts (eV) and is the energy of the precursor ions within the isolation window immediately before they collide with the collision gas and undergo fragmentation to form product ions [54]. MS2 experiments often include multiple analyses using the same isolation window at different collision energies because fragmentation patterns change as the energy increases.

Ion Mobility Spectrometry

Ion mobility (IM) spectrometry is an emerging approach for separating and identifying ions based on their gas-phase mobility, often reported through collision cross section (CCS) values [55-58]. Trapped ion mobility spectrometry (TIMS), a common IM technique in MSI, traps ions in an electric field against a counter-flowing gas and separates them according to their mobility as they are released by ramping the electric field under controlled gas flow. IM can help distinguish isomers and isobars that overlap in MS1 [59].

Ion mobility spectrometry integrated with mass spectrometry imaging (IM-MSI) has become a powerful approach for improving metabolite annotation and acquiring additional structural information in biological tissues, supported by instrument platforms such as the timsTOF (Bruker Daltonics). In these systems, ions are separated on the millisecond timescale based on their gas-phase mobility before rapid TOF mass analysis, enabling efficient acquisition of

mobility-resolved MS1 (and if necessary, MS2) data. By providing information on two complementary descriptors (i.e., m/z and ion mobility), IM-MSI improves the separation of interferences and strengthens direct metabolite identification in tissue samples.

Data collection and analysis

Mass spectrometry imaging experiments can generate terabytes upon terabytes of data: individual spectra from certain experiments may contain millions of data points [20]. The most basic types of MSI data comprise spot coordinates for each location in the sample that was analyzed together with m/z and intensity arrays for each spot. However, the increasingly common use of ion mobility methods in MSI means that the data may also include ion mobility arrays for each spot. MSI software must therefore be able to process large amounts of data quickly and reliably.

MSI software

The first software for processing MSI data, Biomap [60, 61], was developed in 2002 by Dr Markus Stöckli at Novartis. It was succeeded by flexImaging [62, 63] (Bruker Daltonics) in 2005 and ImageQuest [63, 64] (Thermo Fisher Scientific) in 2007. Several other proprietary and open-source software packages for MSI have since been developed.

The manufacturer of the instrument used in a given experiment often dictates the software used for initial analysis because raw data files are typically instrument-specific. For instance, mass spectrometers made by Bruker Daltonics are supported by SCiLS Lab, which enables users to visualize data, perform feature selection, export results, and conduct statistical analyses. This software is regularly updated to meet user demands. Similarly, other instrument vendors such as Waters [65], Shimadzu [66], Thermo Fisher Scientific [67], and MassTech [68] promote their own proprietary software for data acquisition and analysis. This reliance on vendor-specific tools in the initial processing step contributes to the diversity of MSI workflows.

Despite the increasing openness of software in other omics fields, including mass spectrometry, MSI remains largely dependent on proprietary instrument vendor formats. Although the open file format imzML [69] exists, raw data files are still predominantly proprietary, often requiring specialized conversion software for broader compatibility. SCiLS Lab offers multi-vendor support for converting MSI data, and the free software package msConvert [70], provides partial compatibility with certain raw data formats, despite not being

specifically designed for MSI. These data format issues complicate the development of MSI software pipelines, necessitating careful consideration of compatibility and standardization to improve workflow flexibility.

Annotation and identification

In an MSI workflow, the initial processing of the raw data is followed by the annotation and identification of relevant m/z features. Several open-source software tools are actively being developed for this purpose (Table 1). MSI annotation tools have traditionally focused on assigning molecular candidates on the basis of accurate mass data and knowledge of expected ion species, which are typically the protonated or metal ion adducts formed when using conventional MALDI matrices or other ionization techniques. However, the introduction of reactive derivatizing matrices, such as FMP-10 [38], has created a need for specialized annotation tools that can account for derivatization products and the more complex ion chemistry of derivatization-based workflows.

Table 1 lists several notable open-source software tools for MSI feature annotation and signal filtering. These tools include desktop, script-based, and cloud implementations, and typically rely on accurate-mass matching to expected ion species or MSI-specific quality control data to flag matrix- and off-sample signals. Dedicated support for derivatization-driven ion chemistry remains relatively uncommon. The following section summarizes the principles of metabolite identification and complementary evidence that can support initial identifications, including MS2 and ion mobility data.

Metabolite Annotation and Identification

Metabolite annotation and identification [59] are crucial for interpreting mass spectrometry results. Annotation typically involves matching observed peak m/z ratio values against entries in databases such as the Human Metabolome Database (HMDB) [71]. For conventional matrices, this process is simplified by calculating the mass, derived by subtracting the adduct mass from the observed m/z ratio. However, derivatizing matrices require additional processing steps for accurate annotation of m/z lists. Paper I introduces Met-ID, a software tool designed for automated annotations of feature lists derived from mass spectrometry experiments, specifically addressing the complexities associated with derivatizing matrices.

Table 1. Examples of open-source metabolite annotation software for metabolite annotation and related signal filtering in MSI. Tools are categorized by their primary target class, whether they explicitly support derivatization-based workflows, their implementation (desktop, script or cloud), and whether they were developed specifically for MSI. Most tools focus on conventional adduct/isotope annotation or removal of matrix/off-sample signals, whereas Met-ID is tailored to reactive-matrix derivatization data and runs locally without requiring cloud upload. The functionality of the Met-ID software (Paper I and II) is described in the Results and Discussion section.

Open MSI Software [reference]	Target	Derivatization	Type	MSI specific
Met-ID [72]	Metabolites, lipids, small molecules	Yes	Desktop	No
Alex123 [73]	Lipids	No	Desktop	No
Cyclobranch 2 [74]	Peptides, Small molecules	No	Desktop	No
HIT-MAP [75]	Peptides, proteins	No	R script	Yes
Mass2adduct [76]	Alkali metal adducts, matrix adducts, isotopes	No	R script	Yes
massPix [77]	Lipids	No	R script	Yes
OffsampleAI [78]	Off sample peaks	No	Cloud	Yes
rMSIannotation [79]	Metabolites, peptides	No	R script	Yes
rMSIcleanup [80]	Matrix signals	No	R script	Yes
Metaspace [81]	Metabolites, lipids, small molecules	No	Cloud	Yes

MS2 data are commonly used when MS1 data alone are insufficient to determine a peak’s identity, as described above. MS2 spectra provide structural information by revealing product ions that facilitate identification of the parent compound. Additionally, MS2 spectra can be compared to reference spectra of chemical standards, although most such spectra were generated using conventional matrices. There is currently no universally accepted method for assessing spectral similarity, but several approaches have been proposed. One commonly used method, cosine similarity [82, 83], treats the spectra as vectors, and evaluates their similarity by calculating the cosine of the angle between them, yielding a score ranging from -1 (perfect dissimilarity) to 1 (perfect similarity). However, since intensities cannot be negative the MS2 similarity scores range from 0 (no similarity) to 1 (perfect similarity).

By resolving structural differences that may not be distinguishable using m/z ratios alone, IM can effectively differentiate isomers, providing an additional dimension of separation that enhances molecular identification in complex samples. IM can also be integrated with MS2 to further improve

structural elucidation, making it a powerful tool for analyzing biomolecules, metabolites, and other structurally similar compounds.

Collision Cross Section

Ion mobility is often reported in terms of the collision cross section (CCS) values, a difficult to estimate property of biological molecules. The CCS values relevant to IM are momentum transfer cross sections rather than strict collision cross sections because they do not depend solely on whether the analyte collides with the collision gas. Traditional hard-sphere collision models [84] are therefore inappropriate for computing the mobility derived cross sections required in IM [85]. Numerous attempts have been made to estimate and predict CCS for IM experiments using various machine learning and deep learning models (table 2) with differing applications, results, and limitations.

A major obstacle to the use of machine-learning models for CCS prediction is a lack of publicly available training data, which can limit generalizability and may lead to overfitting. Good model design can minimize these problems, but de Cripian et al [86] showed that leading methods struggle to predict data previously unseen by the model.

Table 2. Recently published software and models for predicting collision cross section (CCS) values. The listed tools use machine-learning and deep learning strategies to estimate CCS values for metabolites and other small molecules to support ion mobility-enabled annotation and identification. The approaches differ in molecular representation (e.g., descriptors/fingerprints, SMILES-derived-features, or graph neural networks), model architecture (e.g., SVR vs deep learning), and the ion mobility modality and calibration used to generate training labels.

Author [reference]	Name	Year
Plante [87]	DeepCCS	2019
Yang [88]	Not named	2022
Zhang [89]	AllCCS2	2023
Guo [90]	SigmaCCS	2023
Wisnaitayakorn [91]	SVR_CCSPrediction	2024
Hart [92]	Mol2CCS	2024
Xie [93]	GraphCCS	2024
Bouwmeester [94]	CCSHybrid	2024
Bjärterot [Paper II]	CCSSim	2026

Multimodality

Recent advances in spatial omics technologies have enabled the simultaneous investigation of multiple molecular classes in intact tissue sections, providing unprecedented insights into complex biological systems. Among these approaches, MSI and SRT have emerged as powerful techniques for mapping the spatial distribution of metabolites, lipids, and proteins, alongside gene expression at a genome-wide scale. MSI facilitates the untargeted detection of biomolecules by measuring their m/z ratios directly from tissue sections, while SRT captures spatially resolved gene expression by sequencing mRNA transcripts in their native tissue context.

Pathway analysis [95, 96] is a well-established method for investigating molecular interactions within biological systems, integrating information from genes, proteins and metabolites to infer functional relationships. However, traditional pathway analysis often focuses on individual molecular layers in isolation, limiting the ability to capture the full complexity of biological processes. The integration of MSI and SRT within the same tissue section provides a multimodal framework that enhances the spatial resolution of molecular interactions, enabling the correlation of gene expression patterns with metabolic and lipidomic profiles.

Although multimodal spatial omics is highly promising, important methodological barriers persist. MSI and SRT differ in sample preparation requirements, detection sensitivity, and achievable spatial resolution, which complicates direct co-registration and joint interpretation. In addition, experimental conditions must be carefully balanced to protect RNA quality without compromising MSI signal quality and coverage. For this reason, many multiomics workflows have traditionally used adjacent serial sections to separate the assays. However, this strategy introduces uncertainty in spatial correspondence and limits pixel-to-spot comparisons. More recently, protocols enabling MSI and SRT on the same tissue section have been reported, improving spatial fidelity and enabling more direct molecular concordance. Paper II, describes the combined application of MSI and SRT to a single section and details the methodological adaptations required for robust, spatially resolved multimodal profiling.

Scientific Programming

Python [11] is a high-level programming language that is widely used in data science and machine learning due to its accessibility and the availability of high-performance libraries such as NumPy [97] and SciPy [98] for computationally intensive tasks. While beginner-friendly and prioritizing readability,

Python sacrifices some raw execution speed by using a garbage collector. In contrast, lower-level languages such as C and C++ offer higher execution speeds and multiprocessing, which makes them preferable for performance-critical systems, at the cost of complex memory management. The development of Met-ID, which is a centerpiece of this thesis, called for both high performance and robust memory safety. These requirements prompted the use of Rust [99], a modern language designed to address these combined requirements without compromising execution speed. The following sections explain the specific advantages of the different programming languages and frameworks that were used when developing the projects presented in this thesis.

Python

Python has long been the dominant programming language in data science, serving as the primary language for most machine learning and AI applications despite efforts to expand the range of languages used for deep learning [100]. Created by Guido Van Rossum in 1991, Python is written in C. This allows computationally demanding Python libraries to be written in C to enable fast execution, which is particularly valuable for neural networks and AI applications. However, Python uses garbage collection to automatically clean up unused memory instead of relying on manual memory management. While this simplifies memory management for the programmer, it comes at the cost of lower execution speed. Python is beginner-friendly because it is comparatively similar to English, and it is often taught alongside R in bioinformatics courses. The core principles of Python emphasize readability and simplicity, as reflected in guiding statements such as “Beautiful is better than ugly” [101] and “Simple is better than complex” [101]. This focus on code readability has contributed to Python’s widespread adoption, despite the fact that it often comes at the expense of execution speed.

Rust

Rust uses the concept of ownership, which offers stricter control over memory access than C and C++ by default while avoiding Python’s reliance on a garbage collector. In the ownership system, data is assigned at compile time to an owner variable that has sole access to the corresponding memory. Rust allows ownership of a value to be temporarily shared through borrowing, but the original owner never changes unless the value is explicitly moved. The value is automatically dropped when the owner goes out of scope, preventing memory leaks without requiring a garbage collector. This eliminates many of the developer burdens associated with manually allocating and deallocating memory in C and C++.

These features have contributed to Rust's sustained popularity, which is evidenced by its top ranking in the Stack Overflow developer survey over several years [102]. However, the use of Rust in software development also imposes certain limitations, many of which are due to its relative youth: Rust was first introduced in 2012, and its capabilities are still being explored and developed [102]. Notably, there is a significant discrepancy in the availability of libraries between Rust and more established languages such as C++, especially in domains such as bioinformatics and cheminformatics.

Graphical user interfaces

Each of the programming languages discussed above provides mechanisms for developing graphical user interfaces (GUIs). For example, Python offers libraries such as Tkinter and Qt, while GTK is commonly used with C and C++ [103]. In Rust, GUI development can be done using libraries such as Egui [104] and Iced [105]. However, many of these frameworks require the application to be written entirely in a single language, often supplemented by styling languages such as cascading style sheets (CSS) to enhance visual appearance. It is worth noting that these languages are generally not optimized for GUI development to the same extent as JavaScript or TypeScript, languages that are more suitable for creating user interfaces but are less commonly used in data science applications. GUIs enhance software accessibility, particularly for users with limited programming experience. This increased accessibility typically comes at the cost of additional development time compared to command-line interfaces (CLIs).

Tauri

The Tauri framework enables effective collaboration between lower level programming languages such as Rust (which is optimized for memory-safe data analysis) and higher-level languages such as JavaScript and TypeScript that are designed for user interface development and visualization, in combination with HTML and CSS. As such, it facilitates the integration of high-performance systems-level computation with modern, responsive front-end visualization. This prompted its use when developing Met-ID (Paper I & II). In this implementation, Rust was used on the backend to exploit its speed and multi-processing capabilities, while the frontend was developed using TypeScript, HTML, and CSS to produce an interactive and visually intuitive user interface.

Open-source software

Open-source code-sharing platforms such as GitHub [15] and Bitbucket [106] have had a transformative impact on the field of bioinformatics. These platforms enable researchers to host code, software tools, and computational

pipelines at no cost, with the option to maintain repositories privately during development and make them publicly accessible upon completion. This model promotes transparency, reproducibility, and collaborative development both during and after a project's active phase, while also enabling technical details to be viewed under the hood. Given the rapid pace of scientific research, developers often move on to new projects upon completion, leaving limited time for continued software maintenance. Open-source sharing mitigates this by allowing others in the community to access, modify, and extend existing codebases. Such contributions can lead to improvements or customizations without altering the original repository, and beneficial changes can be proposed for integration into the main project via mechanisms such as pull requests.

Continuous integration and deployment

Continuous integration (CI) and continuous deployment (CD) are widely adopted practices for automating software development and release processes. CI/CD pipelines streamline workflows by automating tasks such as building, testing, and packaging software, thereby reducing the need for manual intervention and minimizing the risk of human error. These pipelines are particularly valuable in large-scale software projects targeting multiple platforms, where the complexity of the build process can pose significant challenges. In the development of Met-ID (Paper I & II), a CI/CD pipeline was implemented using GitHub Actions to automatically generate platform-specific executables for Windows, macOS, and Linux in response to codebase updates. This automation enhanced the reproducibility and efficiency of the deployment process.

Artificial Intelligence in Computational Mass Spectrometry

The integration of artificial intelligence (AI) in software development has facilitated the rapid prototyping and debugging of scientific applications, including those used in MSI. When developing Met-ID (Paper I & II) using Tauri, AI tools including ChatGPT were employed to assist in generating code, debugging errors, and optimizing performance. Rust's emphasis on memory safety and performance makes it well-suited for computationally intensive tasks, but its strict borrowing and ownership rules can pose challenges for developers unfamiliar with its ecosystem.

ChatGPT [1] is particularly useful for accelerating Rust development by providing syntax suggestions, explaining compiler errors, and refactoring code for better efficiency. It can assist in handling data processing, multithreading, and file parsing, all essential for analyzing MSI datasets. However,

as Rust has a steeper learning curve than dynamically typed languages like Python, AI-generated code often requires manual refinement to ensure correctness and adherence to idiomatic Rust practices. A key limitation of using AI for Rust programming is that AI models are trained on publicly available information from the internet, meaning that there is less Rust training data than there is for older and more widely used languages. AI-generated Rust code may thus contain logical errors, misuse borrowing rules, or simply make up libraries, preventing programs using such code from compiling. Developers must carefully validate AI-generated code using Rust's strong compiler diagnostics, official documentation, and rigorous testing to ensure reliability. Despite these limitations, AI can significantly improve productivity by automating repetitive coding tasks, accelerating debugging, and suggesting optimizations. When used alongside manual review and testing, it is a valuable tool for developing of high-performance software such as Met-ID, helping researchers efficiently process and analyze complex mass spectrometry data.

Aims

This PhD project focuses on the development and application of computational tools for MSI and spatial multimodal analysis, with Met-ID being its central software contribution. A major challenge in MSI is that it generates vast quantities of data requiring extensive processing, much of which is typically done manually. This project aims to address the computational gap in MSI analysis by developing software and workflows that enable more automated, efficient, and confident metabolite annotation and identification, particularly for derivatization-based MS; and to demonstrate the utility of these tools in multimodal and sample-limited applications. The specific aims of the studies included in this thesis (both published and in progress) are to:

- Develop Met-ID, an open-source GUI for efficient annotation of MSI data, with a particular focus on reactive derivatizing matrices, including MS2 library support and spectral-matching functionality to strengthen metabolite identification. This software is a cornerstone of the PhD project, supporting the automation of MSI data analysis. (Paper I)
- Further develop Met-ID to broaden its applicability and increase confidence in its output by adding support for the open data format imzML, a REST-API for programmatic access, and adding additional modules for improved confidence in annotation by using novel CCS prediction and mixture model scoring procedures. (Paper II)
- Develop and validate a workflow for combining MSI and spatially resolved transcriptomics (SRT) on the same tissue section, including evaluating MSI performance on barcoded Visium gene-expression slides relative to conventional ITO-coated slides. (Paper III)
- Demonstrate sample-efficient metabolomic profiling using MSI by analyzing 1 μ L of crude cerebrospinal fluid (CSF) and determine whether this approach with computational annotation can distinguish de novo Parkinson's disease patients from controls. (Paper IV)

Methods

Human Samples

The human tissue used in Paper III was a postmortem sample from a 94-year-old man diagnosed with Parkinson's disease in Braak stage 3; the time between the patient's death and freezing of the tissue was 9.25 hours. The sample was obtained from the Harvard Brain Tissue Resource Center at the McLean Hospital (Belmont, MA). Analyses were approved by the local ethical committee (Karolinska Institutet, Stockholm, Sweden, 2014/1366-31). All experiments were performed in compliance with all relevant ethical regulations [107].

The CSF samples used in Paper IV were obtained from Karolinska Institutet (Stockholm, Sweden). All investigations were performed in accordance with the Helsinki Declaration and with permission from the hospital's local ethics committee (no. 2019-04967).

Animal Samples Experiments

Murine samples (paper III), were obtained from four 8-week-old male mice (C57Bl/6J) housed in controlled temperature (20°C) and humidity (53%) conditions with cycles of 12 hours light/dark. The mice had a standardized diet of food pellets and water *ad libitum*. All animal work was performed in accordance with the European Council Directive (86/609/EE) and approved by the local Animal Ethics Committee (Stockholms Norra Djurförsöksetiska Nämnd, 3218-2022) [107].

Of the four mice, one served as an untreated control, and three underwent unilateral 6-OHDA lesioning. The lesioned mice were anesthetized with isoflurane (Apoteket AB, Solna, Sweden), pre-treated by intraperitoneal administration of desipramine (25 mg/kg; Sigma-Aldrich, St Louis, MO, USA) and pargyline (5 mg/kg; Sigma-Aldrich), and placed in a stereotaxic frame. Each mouse was injected over 2 min with 3 µg of 6-OHDA in 0.01% ascorbate (Sigma-Aldrich) into the median forebrain bundle of the right hemisphere. Injections were made at -1.1 mm anterior-posterior, -1.1 mm medial-lateral, and -4.8 mm dorsal-ventral relative to bregma and the dural surface.

Postoperatively, buprenorphine was administered subcutaneously for two days for analgesia. Lesions were validated two weeks after 6-OHDA administration by injecting apomorphine (1 mg/kg) and assessing rotational behaviour, which correlates with lesion severity. After euthanasia, brains were snap-frozen for 3 s in dry ice-cooled isopentane and stored at -80°C

MSI workflow

Sample preparation and data collection

An MSI experiment begins with sample preparation, which involves cryo-sectioning tissue specimens and mounting them onto indium tin oxide (ITO)-coated glass slides or, in the case of biofluids, spotting them onto MALDI target plates. In MALDI-MSI, a matrix is uniformly applied to the sample surface to facilitate desorption/ionization during analysis. The choice of MALDI matrix is experiment-dependent and is made on the basis of the chemical properties of the target biomolecules. In this work, the matrix was sprayed onto the tissue samples using a TM-Sprayer (HTX Technologies, Chapel Hill NC, USA). After matrix application, the prepared samples were introduced into the mass spectrometer for analysis. The projects presented in this thesis primarily utilized an FT-ICR (Solarix 7T 2-omega, Bruker Daltonics, Bremen, Germany) mass spectrometer, which offers high mass accuracy and resolution for molecular analysis. Additional experiments were conducted using a MALDI-2 timsTOF fleX mass spectrometer (Bruker Daltonics), which offers higher acquisition speed and improved lateral resolution along with IM.

Data analysis

Following data acquisition, raw MSI data were imported into SCiLS Lab (Bruker Daltonics) for feature detection and statistical analysis. SCiLS Lab provides tools for peak picking, intensity normalization and spatial segmentation. Selected features were then exported and annotated using Met-ID (Paper I) by specifying the MALDI matrix used in the experiments, enabling precise metabolite annotation. Further statistical analyses were performed in Python using Jupyter notebooks, incorporating libraries for data processing, visualization, and multivariate analysis. Additionally, SIMCA (Umetrics Sartorius, Umeå, Sweden) was used to perform partial least squares discriminant analysis (PLS-DA), enabling classification and biomarker discovery based on MSI-derived spectral patterns (Paper IV).

Spatial Multimodal Analysis Workflow

For the project on integrating spatial analyses within the same tissue section (Paper III), the samples were cryo-sectioned and mounted onto non-charged, barcoded gene expression arrays used in SRT, rather than the ITO glass slides typically used for MSI. MSI analysis was performed as outlined above, with the key difference being that the matrices were washed off after MSI acquisition. After MSI, the samples were transferred on dry ice to the SRT laboratory in Stockholm, where they were stained with a hematoxylin and eosin (H&E) stain and imaged with bright field microscopy. Finally, SRT analysis was performed to map gene expression within the tissue samples, enabling multimodal spatial analysis combining metabolomic and transcriptomic data by aligning the two modalities computationally.

Spatial alignment of MALDI-MSI and SRT data from the same tissue section

Co-registration of the two modalities from the same tissue section was performed in R using the `ManualAlignImages` function within `STUtility`'s interactive Shiny interface, with manual alignment based on tissue images. Because only H&E images were available for the transcriptomics modality, the workflow was adapted such that MSI sampling coordinates were aligned directly to the corresponding H&E image, rather than relying on an MSI-specific optical image. Before alignment, MSI data were subjected to PCA to identify and remove acquisition points outside the tissue section. After manual alignment, cross-modality spot matching was performed by assigning each MSI data point to its nearest SRT spot using `k`-nearest neighbors (`dbscan`; $k = 5$). Non-overlapping regions were removed by excluding MSI–SRT pairs with distances >35 pixels (threshold chosen empirically based on the neighbor-distance distribution). Given the lower spatial sampling density of MSI relative to Visium, neighbor selection was performed from MSI to SRT; the resulting matched pairs were used to subset both datasets and generate new Seurat objects with the two aligned modalities stored as separate assays.

Spatial cross-modality correlation and gene–metabolite association analyses

For targeted cross-modality associations, spatial correlations between a metabolite of interest (e.g., dopamine intensity) and gene expression were computed across the aligned MSI–SRT pairs using pairwise Pearson correlations between the metabolite signal and each gene. Multiple-testing correction was applied using the Benjamini–Hochberg procedure, retaining associations at $FDR < 0.01$.

To more broadly link metabolite peaks to transcriptional programs, peak-associated gene ranking was performed for the 2,000 most highly variable genes using k-nearest-neighbor graph methods, followed by joint PCA of the two modalities within anatomical groupings (e.g., striatum vs substantia nigra). A binary search tree was built on the PCA co-embedding of the SRT data using cosine similarity, and for each MSI peak the ten nearest neighbors in the embedded space were retrieved to construct an integrated gene–peak neighborhood graph. This graph was then used to identify spatially co-detected gene/peak modules via Spinglass community detection and was visualized using a weighted Distributed Recursive Layout.

Software and Programming languages

Software

Software used for MSI data acquisition and processing included *ftmsControl* for instrument control, *flexImaging* for visualization and basic analysis, and *SCiLS Lab* (all from Bruker Daltonics) for feature detection and generation of *m/z* feature lists and ion images. *Met-ID* was used for feature annotation, enabling identification of derivatized metabolites. SRT data were pre-processed using *Space Ranger* (10X genomics, Stockholm, Sweden) and visualized with *ImageJ* (National Institutes of Health, Bethesda, USA). *Loupe Browser* (10X Genomics) was used to interactively explore and analyze SRT datasets. Image alignment between MSI and SRT was done manually in Python by removing off-tissue pixels identified by PCA, and using affine transformations to align the MSI image to the H&E image as the affine transformation between H&E and SRT was established during the experiment.

Correlations between MSI on ITO glass slides and MSI for SRT on Visium glass slides were evaluated using Python after selecting a reference (ITO-derived) feature list in *SCiLS* and exporting the data through the API.

Programming languages

The development of code and software across the projects utilized several programming languages, including Python, Rust, and TypeScript, along with HTML and CSS for user interface design. R was also employed for spatial transcriptomics analyses, but not for MSI applications. Multiple versions of Python 3 (3.7–3.10) were used to analyze MSI data, develop analysis workflows, and implement early versions of *Met-ID*. Key Python packages included *RDKit* [108], *NumPy* [97], *SciPy* [98] and *scikit-learn* [109], among other widely used scientific computing libraries. *Jupyter Notebooks* [110] were used extensively for developing and visualizing data analysis scripts.

Met-ID was developed using the Tauri framework (v2), with Rust serving as the backend programming language and TypeScript, HTML, and CSS supporting the frontend interface. The Rust toolchain (cargo version 1.90) was used alongside widely adopted packages such as `serde` [111], `rayon` [112], and `rusqlite` [113]. On the frontend, TypeScript (version 5.4.5) was used in combination with `D3.js` [114] for interactive data visualizations and `smiles-drawer.js` [115] (version 2.1.5) for rendering chemical structures.

Results and Discussion

The Met-ID Software (Paper I & II)

The development of Met-ID was a central component of this thesis and was undertaken to automate and streamline key steps in MSI data analysis. Met-ID allows users to input lists of m/z features for rapid annotation, while accounting for the specific matrix chemistry employed in the experiment (Figure 3). Unlike other MSI annotation tools, Met-ID is designed to support workflows using derivatizing (reactive) matrices that generate fundamentally different ion species to those produced with conventional matrices.

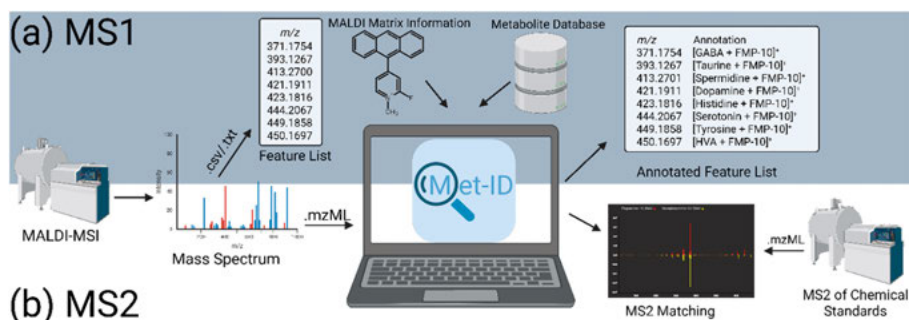


Figure 3. Overview of Met-ID workflows. (a) In the MS1 workflow, Met-ID imports a feature list from a .csv or .txt file, along with information on the MALDI matrix chemistry and a pre-compiled metabolite database, resulting in a list of annotated features for downstream analysis. (b) In the MS2 workflow, Met-ID imports MS2 spectra as .mzML files and compares experimental spectra to reference spectra in its database, enabling annotation and identification based on MS2 evidence as well as MS1 data.

Derivatizing matrices such as FMP-10 carry a permanent positive charge and form covalent bonds with target metabolites. Specifically, FMP-10 undergoes a nucleophilic substitution reaction in which a primary amine or phenolic hydroxyl group on the target metabolite reacts with a fluorinated carbon on FMP-10. This reaction forms a stable covalent bond between the derivatizing matrix and the metabolite. The resulting derivate is positively charged and incorporates a chromophore, enhancing its ability to absorb MALDI laser energy and thus facilitating efficient desorption and ionization from the tissue surface.

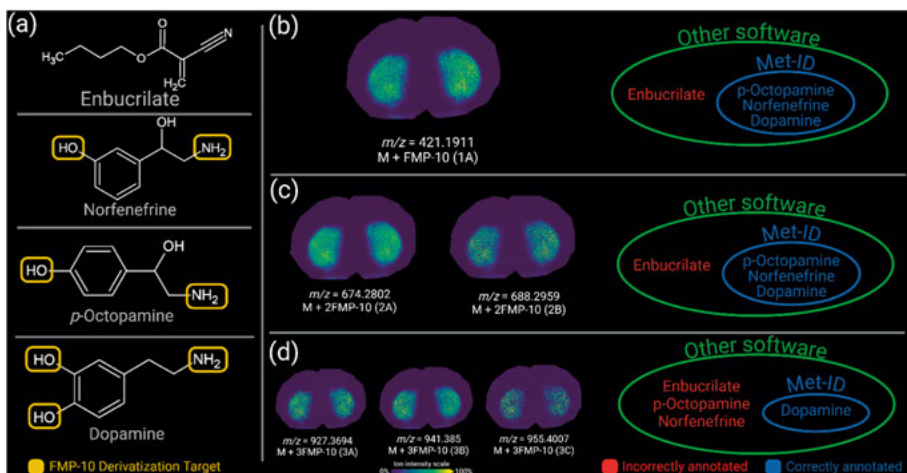


Figure 4. Comparison of Met-ID to conventional MSI annotation approaches for derivatized species. The four metabolites shown in (a) have the same molecular formula ($C_8H_{11}NO_2$) and monoisotopic mass (153.07898). Panels (b-d) illustrate differences in annotations for (b) single, (c) double, and (d) triple derivatized ions. Unlike other annotation software, Met-ID incorporates derivatizing-matrix chemistry and applies structure-based rules to exclude implausible candidates. This example involves the reactive matrix FMP-10, which derivatizes primary amines and phenolic hydroxyl groups, (highlighted in yellow in (a)). Of the four isomeric metabolites, dopamine alone contains three such groups, so Met-ID identifies it as the only one that could produce the triple derivatized peaks; tools that do not account for molecular structure, would annotate all four isomers. The consistent lateral distribution of the annotated ions across the tissue section (b-d) further supports the assignment to dopamine. Ion images are shown using a rainbow scale (representing ion intensity) for clarity. Lateral resolution: 100 μm .

The use of derivatizing MALDI matrices such as FMP-10 adds complexity to metabolite annotation due to the mass shift introduced upon covalent modification of the target molecules. In addition, FMP-10 can react multiple times with a single metabolite, and may undergo side reactions such as the loss of a methyl group, generating multiple distinct peaks corresponding to the same compound. This can facilitate discrimination between analytes that would otherwise be difficult to distinguish, such as the four isomers depicted in Figure 4: although they have the same molecular formula and monoisotopic mass, they form different numbers of products after FMP-10 derivatization, giving rise to the distinct peak ion images shown in Figures 4b–d. Conventional annotation software lacks structural filtering capabilities and would thus annotate all four isomers across all six peaks, whereas Met-ID applies derivatization-specific structural logic to restrict annotations only to chemically plausible candidates. For example, enbucrilate is not expected to be derivatized by FMP-10, whereas dopamine can undergo triple derivatization.

Met-ID also incorporates a curated database of FMP-10–derivatized chemical standards analyzed using MS2. This lets users compare experimental data to known reference spectra to support confident metabolite identification. Spectral matching in Met-ID is performed using cosine similarity, with user-defined binning parameters. Additional visualization features include mirrored spectrum comparison and the ability to export plots either as image files or in CSV format for use with third-party plotting tools.

Finally, Met-ID offers a name-based search functionality, letting users query the database using the names of target metabolites rather than m/z values. This feature returns the expected derivatized mass for each supported matrix, enhancing the flexibility and accessibility of the annotation process.

Development of Met-ID

A key principle in the development of Met-ID was to emphasize personalization and customization. While the software’s default configuration is designed to suit a broad range of typical use cases, it was important to let users tailor the tool to their specific research needs. This design choice was motivated by the inherent difficulty of anticipating every potential application in scientific research, and also by a desire to enable user-level modifications without requiring programming expertise.

Initial versions of Met-ID were implemented in Python and had functionality similar to that of the current version for discrete tasks. However, Python’s limited support for parallel processing proved to be a major constraint during development. This limitation arises because Python’s Global Interpreter Lock (GIL) restricts execution to a single thread within the interpreter, making it difficult to implement efficient multiprocessing without interfacing with lower-level languages. Consequently, Met-ID was reimplemented using the Tauri framework and Rust, which offers robust support for concurrency and improved computational performance. Although this transition temporarily slowed development, it ultimately led to significant improvements in execution speed and enabled the development of additional high-performance tools within the project.

The Met-ID metabolite identification algorithm

Met-ID is built around an SQLite3 database containing various tables with information on metabolites, matrices, and functional groups. The identification algorithm uses this database in two main steps.

SQL Query

The first step is the SQL query. The user defines the list of m/z features to query as well as the type of metabolites and tissue. If querying a conventional matrix, the allowed adducts can be selected and if querying a derivatizing matrix, the functional groups to allow can be selected. Using these parameters, Met-ID can construct an SQL query to only allow matching of plausible adducts or derivatives. The SQL query will check whether a target metabolite contains the selected functional groups if querying a derivatizing matrix, and will also determine how many such groups are present if the matrix permits multiple derivatizations. If a metabolite contains one or more of the selected functional groups, the mass of the derivatized product must be between the lowest and the highest m/z in the m/z list. If these conditions are satisfied, the execution of the SQL query will generate these adducts or derivatives.

Mass Matching

The second step is mass matching: the list of metabolites from the database is sorted by mass and mass matching is performed using the `bisect_left` algorithm. This simple bisection algorithm bisects the data (i.e., splits in half) and determines which side of the bisection contains the query number. That side is then bisected again, and the process is repeated until the index of the number is determined. The complexity of the bisection algorithm is $O(\log n)$, so it is scalable and is not significantly impacted by increasing sample size. During mass matching, the bisection algorithm is employed twice for each m/z to find all metabolites within a user specified distance (e.g., 2ppm) of the target feature. Values within this range are returned as annotations.

Met-ID mass calibration

The mass calibration algorithm takes in known peaks, using paired theoretical and observed m/z values as inputs and fits a curve to them in Rust using `nlopt`'s [116] `Newuoa` [117] and `Bobyqa` [118] algorithms consecutively. The masses are recalibrated in the annotation algorithm according to equation 1 below:

Equation 1

$$m_a = am_e^2 + bm_e + c$$

Here, m_a is the adjusted m/z , m_e is the experimental m/z , while a , b , and c are the fitted variables

CCS prediction

To improve Met-ID and enable improved annotation of IM-MSI data, the Met-ID database had to be extended to include CCS values. Building a table of CCS values for Met-ID required *in silico* predicted CCS values for both conventional and derivatizing matrices, which necessitated a universal and generalizable prediction method. Met-ID 2.0 therefore introduced a novel prediction model, called CCSSim (Figure 5) that combines molecular simulations and linear regression. In CCSSim, models of database metabolites and their adducts are built in bevy [119], a Rust-based game engine, with the radius of the collision gas (N₂, 2.106 Å) added to each atom to simulate the surface on which the nitrogen gas may collide with the target. The collisions are simulated using “shape casting” and several simulation-based metrics are collected as the simulation progresses. These metrics are then used with other molecular descriptors to build a linear model to predict CCS values for each molecular species. Finally, the predicted CCS values are added to the Met-ID database for use in multimodal annotation.

The CCSSim method achieves a mean absolute percentage error (MAPE) of 2.9% on a training set consisting of publicly available CCSBASE [120] data. CCSSim also outperforms CCSBASE in predicting FMP-10 derivatized chemical standards (MAPE of 2.8% for CCSSim and 8.6% for CCSBASE, respectively) (Paper II).

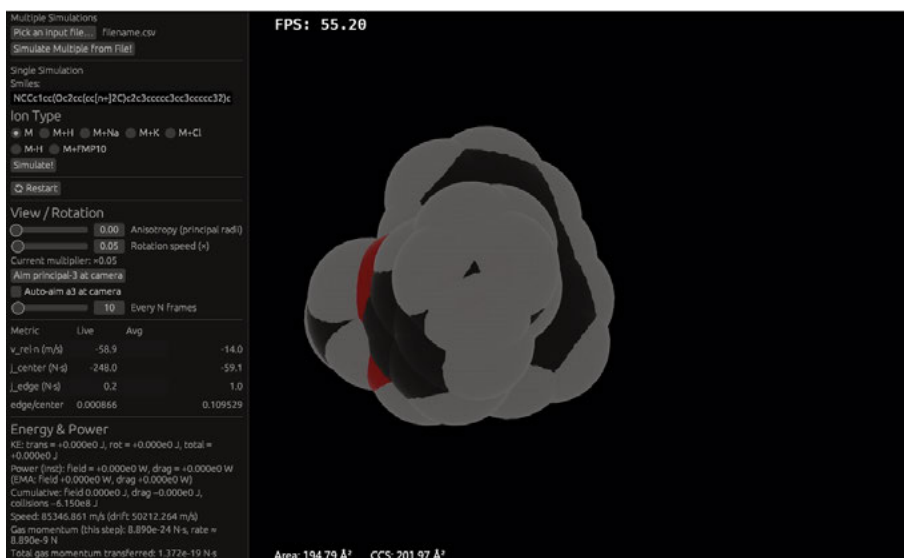


Figure 5. A representative CCSSim simulation of [dopamine + FMP10]⁺. The left window contains the user interface, with fields to input SMILES strings, buttons for importing files for multiple sequential simulations, sliders for tuning the simulation by adjusting the molecule’s rotation, as well as some performance metrics. The remaining area shows the simulation (in this case, FMP-10 derivatized dopamine), with the current area and current CCS value displayed at the bottom. The simulation frames per second (FPS), i.e., how many times per second the simulation updates, is displayed at the top.

Mixture model for identification confidence

Multimodal annotation of IM-MSI data allows annotation confidence models to be built. Rather than defining targets and decoys using a frequentist approach, which is difficult with ill-defined false positives. Met-ID uses a mixture model based on Bayesian statistics using prior probabilities and likelihoods. One advantage of a mixture model is the ability to assess competing hypotheses as well as the hypothesis that an annotation may not be found in the database. As all databases are finite, there is always a possibility that a feature seen in the data will not exist in the database. Outside of database metabolites, IM-MSI data commonly feature multimers, fragments, and isotopes, many of which will not be found in HMDB. One limitation of the mixture model arises from the statistical power of current state-of-the-art CCS predictions. The mixture model will improve in parallel with predictive models because its specificity depends on the magnitude of the instrumental and predictive errors in the input m/z and CCS data.

The Met-ID REST-API

Met-ID 2.0 has a REST-API that enables high throughput use and facilitates its inclusion in programmatic pipelines. While the Met-ID GUI was designed to let non-programmers customize their pipelines, it was also considered important to support the development of high-throughput pipelines and automation. The Met-ID REST-API is developed for Python and R and may be extended to other languages if necessary. The Met-ID API allows users to add custom metabolites, matrices and functional groups similarly to the GUI as well as running the annotation algorithm.

Validation of Met-ID

Four separate annotation searches were performed using a dataset of 256 m/z features from 36 chemical standards (28 endogenous). Features with m/z values below that of FMP-10 (m/z 268.1121) and peaks corresponding to FMP-10 and its fragments, were manually excluded, refining the dataset to 170 m/z features. This refined dataset was then searched with Met-ID, using four distinct search conditions for comparative evaluation. Specifically, searches were conducted against all HMDB metabolites and against only endogenous metabolites, with and without mass calibration in both cases.

The first search targeted all metabolites that could be derivatized by FMP-10, yielding 95 annotated m/z features (55.9%) including 44 annotations that were matched to derivatives of 35 chemical standards. One standard was not annotated as it fell outside the chosen mass error window of 2 ppm. Multiple derivative peaks were observed for several standards, which was expected given the reactivity of FMP-10. Restricting the search to endogenous metabolites significantly reduced the number of potential candidates from 26,202 to 4,951. Under these conditions, 76 m/z features were annotated, corresponding to 44.7% of the dataset. Of these, 38 annotations (50%) corresponded to derivatives of the chemical standards, covering 28 of the 36 standards included in the dataset. The eight remaining standards were not annotated, because they are not classified as endogenous in the HMDB database.

Met-ID was also used to annotate tissue data, acquired using both FMP-10 and the conventional non-reactive matrix NEDC (in negative mode). The annotation results and the observed improvements following mass recalibration were consistent with those achieved with the chemical standard dataset. These findings demonstrate the applicability of Met-ID in the analysis of biological tissue samples.

Spatial Multimodal Analysis (Paper III)

The aim of paper III was to demonstrate the feasibility of performing MALDI-MSI and SRT on the same tissue section, while maintaining each technique's sensitivity and specificity. Traditionally, studies investigating both gene expression and metabolite distribution have relied on the use of consecutive tissue sections. However, this approach suffers from challenges related to internal validity and spatial alignment. Transitioning from separate analyses of consecutive sections to a combined approach using single sections necessitated the development of protocols that make the two methodologies technically compatible and ensure that the application of one technique does not substantially compromise the other's performance.

Spatial metabolomics and transcriptomics on the same tissue

MSI was performed before SRT because SRT includes washing steps that would be detrimental to the small metabolites studied by MSI. To demonstrate the efficacy of the sequential multimodal analysis (SMA) workflow, it was essential to verify that MSI was not detrimental to the transcripts analyzed by SRT. MALDI-MSI was therefore performed on half of a tissue section, with the other half serving as a control, after which the slide was coated with poly(dT) probes and analyzed by SRT. The transcriptomic data obtained from the two halves did not differ significantly, indicating that the MSI analysis was compatible with subsequent SRT. To further demonstrate the generalizability of SMA, a post-mortem human brain sample was coated with FMP-10 and analyzed using the SMA workflow.

Spatial transcriptomics requires barcoded glass slides, whereas conductive ITO slides are preferred for MSI studies. However, since non-conductive glass slides are also compatible with MSI, tissues were mounted on the primer-coated slides used in the Visium spatial transcriptomics platform. Comparative MSI experiments using these non-conductive glass slides and conventional ITO glass slides revealed strong correlations between the detected peaks ($r > 0.85$ for FMP-10 and $r > 0.74$ for 9AA and DHB) in detected peaks. SMA SRT data also showed a very high correlation ($r > 0.98$) compared to the regular SRT protocol. Figure 6 presents scatter plots and stacked bar plots comparing traditional MSI and MSI performed on barcoded slides for SMA. For both the derivatizing matrix FMP-10 and the conventional matrix DHB, over 90% of detected peaks displayed fold changes of less than $0.5 \log_{10}$ units when comparing the two slide types. Conversely, just over 60% exhibited fold changes in this range when using the conventional matrix 9AA. Most peaks showing greater fold changes had rather low intensities, so their apparent

response to the change of slide type may be influenced by the limit of detection in the analysis.

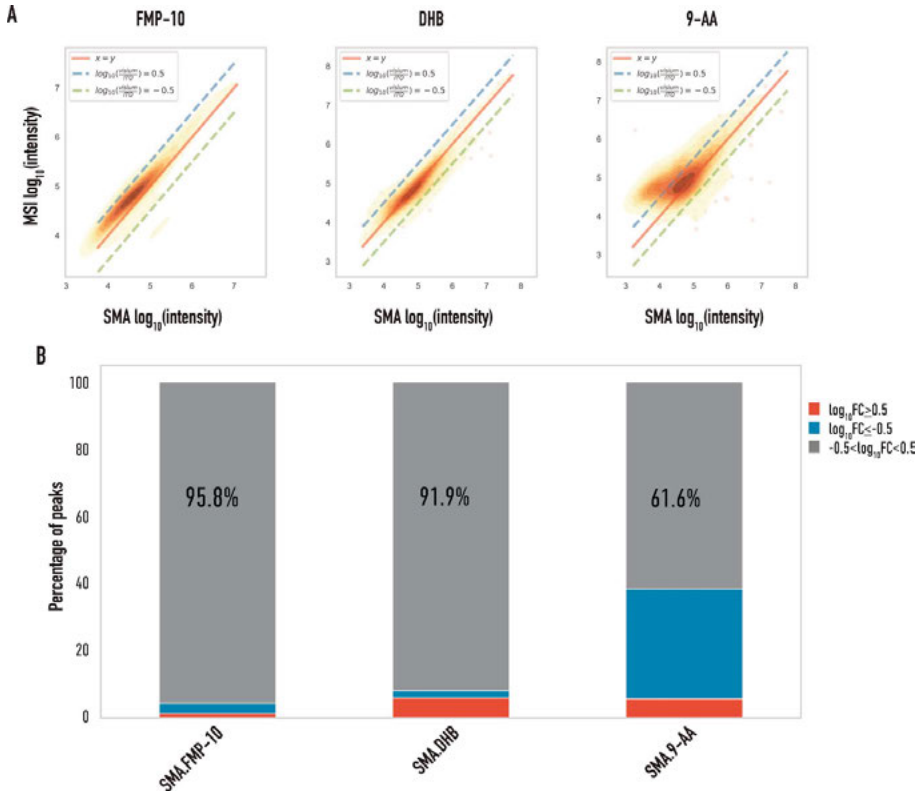


Figure 6. Comparison of SMA-MSI performed on Visium gene-expression slides to standalone MSI on conventional ITO-coated glass. (a) Density scatter plots comparing \log_{10} -transformed peak intensities measured for matched features acquired on Visium slides (SMA-MSI; y-axis) versus ITO-coated glass (standalone MSI; x-axis) for three matrices: FMP-10, DHB, and 9-AA. The solid diagonal line denotes equality ($x = y$). Dashed lines indicate a $\pm 0.5 \log_{10}$ fold-change window, corresponding to intensities within approximately a factor of 3.16 of the ITO measurement. Clustering of points along the diagonal indicates overall agreement in signal intensities between platforms, whereas systematic shifts reflect matrix-dependent intensity differences. (b) Stacked bar plots summarizing, for each matrix, the proportion of detected peaks whose intensities on Visium slides fall within the $\pm 0.5 \log_{10}$ fold-change window relative to ITO (grey) and those that deviate by more than $+0.5$ (red) or -0.5 (blue) \log_{10} fold change. The large grey fractions for FMP-10 (95.8%) and DHB (91.9%) indicate strong agreement between SMA-MSI and standalone MSI, whereas 9-AA shows a lower fraction within the window (61.6%), consistent with greater intensity differences between Visium and ITO for this matrix.

Low-volume CSF Metabolomics for De Novo Parkinson's Disease Classification (Paper IV)

The aim of paper IV was to assess the feasibility of metabolically separating de novo Parkinson's disease (PD) from controls using only one μL of crude cerebrospinal fluid (CSF).

Minimal sample preparation and MALDI-MSI acquisition

Samples were collected, spotted on a MALDI target plate, and sprayed with FMP-10 without further preprocessing. The samples were analyzed on a 7T FTICR (Bruker Daltonics) for high mass accuracy, yielding 2900 m/z features that were selected using SCiLS Lab and annotated with an early version of Met-ID (Paper I & II). This generated annotations for around 300 m/z features.

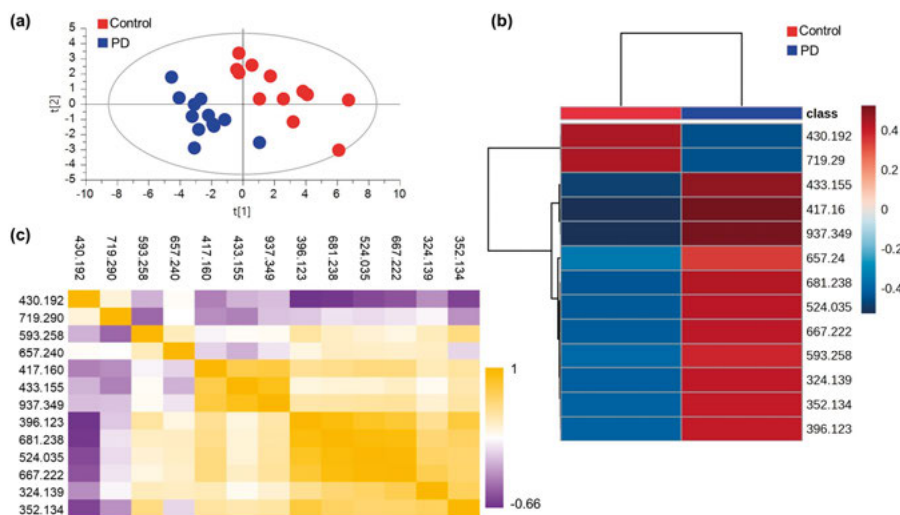


Figure 7. Multivariate modeling of PD-associated metabolic alterations in CSF. (a) PLS-DA score plot of the first two principal components derived after variable selection, showing separation between control and PD samples. (b) Heatmap summarizing the relative contribution/abundance pattern of the selected m/z features across the two groups; negative values indicate lower abundance and positive values higher abundance (according to the color scale). (c) Correlation matrix of the selected m/z features based on Spearman's rank correlation coefficient (r) of the significant m/z values; $P < 0.05$ (two-tailed unpaired t-test). Some identified metabolites are: norcotine (m/z 430.192), hypoxanthine (m/z 657.240), and aminochromes (m/z 417.160 and 937.349, corresponding to single- and triple-derivatized, respectively, and m/z 433.155).

Statistical modeling and feature selection to discriminate de novo Parkinson's disease from controls

Multiple filtering steps were applied to exclude m/z features that were unlikely to represent derivatized metabolites (i.e., features with $m/z < 300$ or high correlation ($r > 0.8$) with known matrix peaks) as well as features associated with technical artifacts and run-order effects. This condensed the feature list to 1,937 m/z features. Principal component analysis (PCA) was performed to assess sample distribution and rule out further outliers, after which a PLS-DA model was constructed. The initial PLS-DA model, based on the full set of 1,937 m/z features exhibited a poor predictive performance ($R^2 = 0.883$, $Q^2 = 0.347$). Variable selection was therefore performed to identify peaks exhibiting significant differences between the two groups, yielding a refined model that included 32 m/z features and showed markedly better predictive performance ($R^2 = 0.862$, $Q^2 = 0.736$; see Figure 7). The model was extensively validated through cross validation ANOVA and permutation testing. Further analyses, including capillary electrophoresis mass spectrometry (CE-MS) for metabolite identification, revealed that the metabolite corresponding to m/z 430.192, whose intensity was significantly reduced in samples from individuals with PD, was norcotinine, a nicotine metabolite. Retroactive examination of the patients' smoking habits showed that the Parkinson's group included a high proportion of smokers; smoking correlates strongly with norcotinine levels in the body.

Conclusions

The ongoing growth in the volume and complexity of MSI data means there is an increasing need for robust high-throughput software to support MSI data analysis. This thesis describes how this need has been addressed through the development and application of Met-ID, a novel annotation software package, together with complementary bioinformatic workflows that demonstrate the utility of MSI in multimodal analysis and in low-volume sample settings.

Paper I presents Met-ID, a software tool for metabolite identification in MALDI-MSI datasets that can rapidly generate candidate metabolite assignments from m/z feature lists. By explicitly incorporating derivatization chemistry and enabling MS2 spectral matching, Met-ID supports structure-informed filtering of candidate annotations. Furthermore, its mass recalibration functionality improves annotation accuracy, reduces false-positive identifications, and strengthens the reliability of downstream analyses.

Paper II extends Met-ID by adding simulation-based CCS value prediction and a mixture model framework for determining posterior probabilities of metabolite annotations. Together with support for open-source data formats such as imzML and programmatic access through a REST-API, these developments advance Met-ID 2.0 toward a more complete open-source platform for MSI annotation and identification.

Paper III demonstrates the feasibility of integrating metabolite imaging and spatial transcriptomics on the same tissue section. This represents an important methodological advance for multimodal spatial omics and provides a foundation for further development of single-section multimodal workflows

Paper IV describes a method for distinguishing de novo Parkinson's disease patients from controls using only 1 μL of crude cerebrospinal fluid (CSF). The approach requires minimal sample preparation and offers advantages over more sample-intensive workflows by minimizing consumption of limited patient materials and reducing analytical time.

Populärvetenskaplig sammanfattning

Denna avhandling behandlar utveckling och tillämpning av metoder för avbildande masspektrometri (MSI) med fokus på småmolekylära analyser i biologiska vävnader och biofluider. MSI möjliggör spatialt upplöst detektion av molekyler genom att registrera masspektrometriska data från definierade positioner i provet.

Matrix-assisterad laserdesorption/ionisering (MALDI) är den mest använda tekniken för MSI. Effektiv jonisering av vissa metaboliter är begränsad med konventionella matriser, vilket har lett till utveckling av derivatiserande (reaktiva) matriser. Dessa bildar kovalenta derivat med utvalda funktionella grupper och kan introducera en permanent laddning, vilket förbättrar detekterbarhet. Derivatisering medför samtidigt ökad analytisk komplexitet genom selektiv reaktivitet och möjlig multipel derivatisering av samma analytsubstans, vilket kan generera flera signaler från en och samma metabolit och därmed komplicera annotering.

För att adressera detta utvecklades Met-ID (arbete I), en programvara för annotering och identifiering av metaboliter i MSI-data med explicit stöd för derivatiseringskemi. Met-ID modellerar matris- och reaktionsspecifika derivat och tillämpar strukturberoende regler för att filtrera kemiskt osannolika kandidatannoteringar, vilket minskar manuellt arbete och förbättrar specificiteten vid derivatiseringsbaserade arbetsflöden. Programvaran integrerar även tandem masspektrometri (MS²)-baserad evidens genom spektral jämförelse mot referensspektra från derivatiserade standarder.

En känd begränsning i MS-baserad metabolomik är separation av isomerer, vilka ofta är isobara i MS¹. Detta hanteras traditionellt med MS², där fragmentmönster ger strukturell information. Som komplement har jonmobilitetspektrometri (IM) etablerats som en ytterligare separationsdimension baserad på gasfasrörlighet, ofta uttryckt som kollisionstvårsnitt (collision cross section, CCS). I arbete II vidareutvecklas Met-ID för IM-baserade dataset genom integration av CCS-information, inklusive prediktion/beräkning av CCS och statistisk modellering för att ge sannolikhetsbaserad viktning av kandidatannoteringar.

Avhandlingen omfattar även multimodala strategier. I arbete III etablerades en metod för att kombinera MSI och spatialt upplöst transkriptomik (SRT) på

samma vävnadssnitt, vilket eliminerar registreringsfel som uppstår vid användning av seriella snitt och möjliggör korrelation mellan metabolitdistribution och genuttryck.

Slutligen demonstreras i arbete IV en minimal provvolymstrategi för biofluidanalys, där 1 μL cerebrospinalvätska analyserades med MALDI-MSI för att särskilja obehandlade Parkinsonpatienter från kontroller. Detta tillvägagångssätt underlättar biomarkörstudier i provbegränsade kliniska material.

Sammanfattningsvis presenterar avhandlingen metodologiska och beräkningsmässiga innovationer som förbättrar automatisering, noggrannhet och multimodal integration inom MSI, med tillämpningar för biomarkörforskning och spatial metabolomik.

Acknowledgements

The work presented in this thesis was supported by the Swedish Research Council (2018-05501, 2018-03320, 2021-03293, 2021-00189, 2022-04198), the Swedish Brain Foundation (FO2021-0318, FO2023-0241, FO2025-0323-HK-250), Science for Life Laboratory, and Uppsala University infrastructure grants. Travel grants for presentations at international conferences were provided by Apotekarsocieteten and the Nordic Metabolomics Society.

Throughout my life and especially during my PhD Studies I have met an incredible amount of people that have all contributed in different ways to the work in this thesis. Words cannot fully describe how grateful I am to all the people listed below, but here is my best effort to do so.

First of all, I want to thank my main supervisor, Per Andrén for giving me the opportunity to come to Uppsala and learn so much about MSI and bioinformatics. You have taught me a lot about working in academia, and how to present my work in a valuable way. I especially appreciate the trust you put in me to try all of my ideas, some of which are featured here in this thesis and some are hidden in a folder far away.

My co-supervisors Anna Nilsson, Reza Shariatgorji and Theodosia Valianatou. The three of you have been an integral part of this project, always giving good guidance and input, answering all of my questions, and most importantly, letting me blabber on about sports during most lunch breaks.

My co-supervisor at KTH, Lukas Käll, for recommending me to the Andrén group. I still remember you said you knew someone in Uppsala working with MSI and looking for a bioinformatician. Your guidance on bioinformatics has immensely elevated the work presented in this thesis.

Ibrahim Kaya and Henrik Lodén for being good office mates and neighbors, and the daily dose of fresh air from going to Hemköp.

I have had the pleasure of working with a lot of great people in the spatial mass spectrometry group, including Per, Anna, Reza, Theodosia, Ibrahim, Henrik, Sofia, Tina, Dominika, Erik, Jordan and Mariya.

The Käll group at KTH, including Gustavo, Markus, Patrick, August, Fei, Yuqi, Alfred, Hilda, Dafni and Jakub. The weekly group meetings have been

invaluable for my development in bioinformatics and I have learned a lot about so many different projects and methods.

Per Svenningsson and his group at Karolinska Institutet for providing high quality samples and collaboration in the projects in this thesis. Joakim Lundberg and his group at KTH, especially Marco Vicari and Reza Mirzazadeh for great collaboration on the SMA project.

The PDD/LBD project, including the Svenningsson group at Karolinska Institutet, the Nilsson group at Stockholm University, the Ayoglu group at KTH and the Odell group at Uppsala University. It has been a great experience being part of such a large collaboration. I have learned a lot about different spatial omics from attending the meetings.

My studies would not have been possible without friends to help me through it, all the way from Jönköping to Stockholm, I have always been very lucky to find really good friends. My IB friends from back home, Alec, Christoffer, Johannes and Mihir. I can't believe it's been almost 15 years since we first met. My friends from Bachelors, including Christoffer, Ellinor, Filip, Graziella, Julia, Linda, Sebastian, Tolis and Veera. I have done a fair bit of calculations throughout this project but even I cannot count all the times we spent in Sebbe's tiny apartment. My friends from the Scilife Master's program including Amparo, Astradeni, Guenther, Kavan, Inés, Josep, Luca, Max, Miren, Sofia and Vijay. I am very happy being your Swedish friend. My quiz team, Felicia, Florian, Joost, Julia, Maria, Mariana and Sebastian. I think other teams were very happy when we weren't there, so they also had a chance to win.

I also want to extend my warmest gratitude to my family and future in-laws. It is hard to summarize years of support in such a short paragraph but I hope you all know how much your support means to me. My future in-laws, Philip, Margaret and Helen, thank you for introducing me to and expanding my knowledge of French wines, English breakfasts and Scottish rugby.

Farmor, Mormor och Morfar, som alltid varit intresserade av den senaste forskningen och speciellt Farmor som kallat mig "sin lille professor" sen jag var ungefär fem år gammal. My brother Erik, who, despite being my harshest critic is also one of my biggest supporters. My guess is I have Justé and Omega to thank for the latter. My parents, who were with me on my first day of school 25 years ago, and have been there for every homework, test and exam - despite not always understanding what I am talking about, they have always supported me through the highs and lows.

Last but not least, Maria, the love of my life, I cannot wait to be the second smartest Dr Bjärterot-White.

References

1. OpenAi. *ChatGPT (GPT-4)*. 2024 2025-04-08; Available from: <https://chatgpt.com>.
2. Dave, T., S.A. Athaluri, and S. Singh, *ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations*. *Front Artif Intell*, 2023. **6**: p. 1169595.
3. Weinstock, M. *Scene at MIT: Margaret Hamilton's Apollo code*. 2016 [cited 2025 2025-12-17]; Available from: <https://news.mit.edu/2016/scene-at-mit-margaret-hamilton-apollo-code-0817>.
4. Abramson, J., et al., *Accurate structure prediction of biomolecular interactions with AlphaFold 3*. *Nature*, 2024. **630**(8016): p. 493-500.
5. Jumper, J., et al., *Highly accurate protein structure prediction with AlphaFold*. *Nature*, 2021. **596**(7873): p. 583-589.
6. Callaway, E., *Chemistry Nobel goes to developers of AlphaFold AI that predicts protein structures*. *Nature*, 2024. **634**(8034): p. 525-526.
7. Altschul, S.F., et al., *Basic local alignment search tool*. *J Mol Biol*, 1990. **215**(3): p. 403-10.
8. Zou, J., et al., *A primer on deep learning in genomics*. *Nat Genet*, 2019. **51**(1): p. 12-18.
9. Eraslan, G., et al., *Deep learning: new computational modelling techniques for genomics*. *Nat Rev Genet*, 2019. **20**(7): p. 389-403.
10. Sen, P., et al., *Deep learning meets metabolomics: a methodological perspective*. *Brief Bioinform*, 2021. **22**(2): p. 1531-1542.
11. Van Rossum, G. and F.L. Drake Jr, *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam.
12. Team, R.C., *R: A Language and Environment for Statistical Computing*. 2021.
13. Bezanson, J.a.E., Alan and Karpinski, Stefan and Shah, Viral B, *Julia: A fresh approach to numerical computing*. *SIAM {R}eview*, 2017. **59**: p. 65--98.
14. Wilkinson, M.D., et al., *The FAIR Guiding Principles for scientific data management and stewardship*. *Scientific Data*, 2016. **3**(1): p. 160018.
15. GitHub. *GitHub*. Available from: <https://github.com>.
16. Microsoft. *Microsoft Copilot*. 2025; Available from: <https://copilot.microsoft.com/>.
17. Caprioli, R.M., T.B. Farmer, and J. Gile, *Molecular imaging of biological samples: localization of peptides and proteins using MALDI-TOF MS*. *Anal Chem*, 1997. **69**(23): p. 4751-60.

18. Norris, J.L. and R.M. Caprioli, *Analysis of tissue specimens by matrix-assisted laser desorption/ionization imaging mass spectrometry in biological and clinical research*. Chem Rev, 2013. **113**(4): p. 2309-42.
19. Korber, A., I.G.M. Anthony, and R.M.A. Heeren, *Mass Spectrometry Imaging*. Anal Chem, 2025. **97**(29): p. 15517-15549.
20. McDonnell, L.A., et al., *Imaging mass spectrometry data reduction: automated feature identification and extraction*. J Am Soc Mass Spectrom, 2010. **21**(12): p. 1969-78.
21. Kallback, P., et al., *A Space Efficient Direct Access Data Compression Approach for Mass Spectrometry Imaging*. Anal Chem, 2018. **90**(6): p. 3676-3682.
22. R. Castaing, G.S., *Microanalyse par Émission Ionique Secondaire*. J. Microscopie, 1962. **1**: p. 395-410.
23. Castaing, R. and G. Slodzian, *Microanalysis using secondary ion emission*. J Mass Spectrom, 2021. **56**(12): p. e4800.
24. Wechsung, R., et al., *LAMMA - a new laser - microprobe - mass - analyzer*. Microsc Acta Suppl, 1978(2): p. 281-96.
25. Borisov, R.S., M.D. Matveeva, and V.G. Zaikin, *Reactive Matrices for Analytical Matrix-Assisted Laser Desorption/Ionization (MALDI) Mass Spectrometry*. Crit Rev Anal Chem, 2023. **53**(5): p. 1027-1043.
26. Calvano, C.D., et al., *MALDI matrices for low molecular weight compounds: an endless story?* Anal Bioanal Chem, 2018. **410**(17): p. 4015-4038.
27. Wang, J., et al., *MALDI-TOF MS imaging of metabolites with a N-(1-naphthyl) ethylenediamine dihydrochloride matrix and its application to colorectal cancer liver metastasis*. Anal Chem, 2015. **87**(1): p. 422-30.
28. Shariatgorji, M., et al., *Direct targeted quantitative molecular imaging of neurotransmitters in brain tissue sections*. Neuron, 2014. **84**(4): p. 697-707.
29. Zhou, Q., A. Fulop, and C. Hopf, *Recent developments of novel matrices and on-tissue chemical derivatization reagents for MALDI-MSI*. Anal Bioanal Chem, 2021. **413**(10): p. 2599-2617.
30. Strupat, K., M. Karas, and F. Hillenkamp, *2,5-Dihydroxybenzoic acid: a new matrix for laser desorption—ionization mass spectrometry*. International Journal of Mass Spectrometry and Ion Processes, 1991. **111**: p. 89-102.
31. Meriaux, C., et al., *Liquid ionic matrixes for MALDI mass spectrometry imaging of lipids*. J Proteomics, 2010. **73**(6): p. 1204-18.
32. Harkin, C., et al., *On-tissue chemical derivatization in mass spectrometry imaging*. Mass Spectrom Rev, 2022. **41**(5): p. 662-694.
33. Beavis, R.C., T. Chaudhary, and B.T. Chait, *α -Cyano-4-hydroxycinnamic acid as a matrix for matrix-assisted laser desorption mass spectrometry*. Organic Mass Spectrometry, 1992. **27**(2): p. 156-158.
34. Vermillion-Salsbury, R.L. and D.M. Hercules, *9-Aminoacridine as a matrix for negative mode matrix-assisted laser desorption/ionization*. Rapid Communications in Mass Spectrometry, 2002. **16**(16): p. 1575-1581.
35. Hou, J., et al., *Organic salt NEDC (N-naphthylethylenediamine dihydrochloride) assisted laser desorption ionization mass spectrometry for identification of metal ions in real samples*. Analyst, 2014. **139**(13): p. 3469-3475.

36. Yang, H., et al., *Streamlined Analysis of Cardiolipins in Prokaryotic and Eukaryotic Samples Using a Norharmane Matrix by MALDI-MSI*. J Am Soc Mass Spectrom, 2020. **31**(12): p. 2495-2502.
37. Mielczarek, P., et al., *Matrix-assisted laser desorption/ionization mass spectrometry imaging sample preparation using wet-interface matrix deposition for lipid analysis*. Rapid Commun Mass Spectrom, 2023. **37**(14): p. e9531.
38. Shariatgorji, M., et al., *Comprehensive mapping of neurotransmitter networks by MALDI-MS imaging*. Nat Methods, 2019. **16**(10): p. 1021-1028.
39. Stübiger, G. and O. Belgacem, *Analysis of Lipids Using 2,4,6-Trihydroxyacetophenone as a Matrix for MALDI Mass Spectrometry*. Analytical Chemistry, 2007. **79**(8): p. 3206-3213.
40. Wheeler, O.H., *The Girard Reagents*. Chemical Reviews, 1962. **62**(3): p. 205-221.
41. Zhou, Q., A. Fulop, and C. Hopf, *Recent developments of novel matrices and on-tissue chemical derivatization reagents for MALDI-MSI*. Anal Bioanal Chem, 2021. **413**(10): p. 2599-2617.
42. Takats, Z., et al., *Mass spectrometry sampling under ambient conditions with desorption electrospray ionization*. Science, 2004. **306**(5695): p. 471-3.
43. Roach, P.J., J. Laskin, and A. Laskin, *Nanospray desorption electrospray ionization: an ambient method for liquid-extraction surface sampling in mass spectrometry*. Analyst, 2010. **135**(9): p. 2233-6.
44. Fletcher, J.S. and J.C. Vickerman, *A new SIMS paradigm for 2D and 3D molecular imaging of bio-systems*. Anal Bioanal Chem, 2010. **396**(1): p. 85-104.
45. Vickerman, J.C., *Molecular imaging and depth profiling by mass spectrometry--SIMS, MALDI or DESI?* Analyst, 2011. **136**(11): p. 2199-217.
46. Thomson, J.J., *XIX. Further experiments on positive rays*. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 1912. **24**(140): p. 209-253.
47. Aston, F.W., *LXXIV. A positive ray spectrograph*. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 1919. **38**(228): p. 707-714.
48. Comisarow, M.B. and A.G. Marshall, *Fourier transform ion cyclotron resonance spectroscopy*. Chemical Physics Letters, 1974. **25**(2): p. 282-283.
49. Marshall, A.G., C.L. Hendrickson, and G.S. Jackson, *Fourier transform ion cyclotron resonance mass spectrometry: a primer*. Mass Spectrom Rev, 1998. **17**(1): p. 1-35.
50. Wiley, W.C. and I.H. McLaren, *Time-of-Flight Mass Spectrometer with Improved Resolution*. Review of Scientific Instruments, 1955. **26**(12): p. 1150-1157.
51. Boesl, U., *Time-of-flight mass spectrometry: Introduction to the basics*. Mass Spectrom Rev, 2017. **36**(1): p. 86-109.
52. Hardman, M. and A.A. Makarov, *Interfacing the orbitrap mass analyzer to an electrospray ion source*. Anal Chem, 2003. **75**(7): p. 1699-705.

53. Makarov, A., *Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis*. *Anal Chem*, 2000. **72**(6): p. 1156-62.
54. Li, B., et al., *Collision-induced dissociation tandem mass spectrometry for structural elucidation of glycans*. *Methods Mol Biol*, 2009. **534**: p. 133-45.
55. Fu, T., et al., *In situ isobaric lipid mapping by MALDI-ion mobility separation-mass spectrometry imaging*. *J Mass Spectrom*, 2020. **55**(9): p. e4531.
56. Michael, J.A., et al., *Mass Spectrometry Imaging of Lipids Using MALDI Coupled with Plasma-Based Post-Ionization on a Trapped Ion Mobility Mass Spectrometer*. *Anal Chem*, 2022. **94**(50): p. 17494-17503.
57. Soltwisch, J., et al., *MALDI-2 on a Trapped Ion Mobility Quadrupole Time-of-Flight Instrument for Rapid Mass Spectrometry Imaging and Ion Mobility Separation of Complex Lipid Profiles*. *Anal Chem*, 2020. **92**(13): p. 8697-8703.
58. McDaniel, E.W., D.W. Martin, and W.S. Barnes, *Drift Tube-Mass Spectrometer for Studies of Low-Energy Ion-Molecule Reactions*. *Review of Scientific Instruments*, 1962. **33**(1): p. 2-7.
59. Djambazova, K.V., et al., *MALDI TIMS IMS of Disialoganglioside Isomers horizontal line GD1a and GD1b in Murine Brain Tissue*. *Anal Chem*, 2023. **95**(2): p. 1176-1183.
60. Robbe, M.F., et al., *Software tools of the Computis European project to process mass spectrometry images*. *Eur J Mass Spectrom (Chichester)*, 2014. **20**(5): p. 351-60.
61. Rauch, M.a.S., M., *Biomap*. 2002, <https://www.ms-imaging.org/biomap/>.
62. Daltonics, B., *flexImaging*. 2005, <https://www.brucker.com/en.html>.
63. Norris, J.L. and R.M. Caprioli, *Analysis of tissue specimens by matrix-assisted laser desorption/ionization imaging mass spectrometry in biological and clinical research*. *Chem Rev*, 2013. **113**(4): p. 2309-42.
64. Scientific, T.F., *ImageQuest*. 2007: <https://www.thermofisher.com/se/en/home.html>.
65. Waters-Corporation. *Waters Corporation homepage*. 2026; Available from: <https://www.waters.com/nextgen/us/en.html>.
66. Shimadzu. *Shimadzu - Excellence in Science*. 2026; Available from: <https://www.shimadzu.com/>.
67. Thermo-Fisher-Scientific, *Thermo Fisher Scientific Inc*. 2026.
68. MassTech. *MassTech*. 2026; Available from: <https://www.apmaldi.com/>.
69. Rompp, A., et al., *imzML: Imaging Mass Spectrometry Markup Language: A common data format for mass spectrometry imaging*. *Methods Mol Biol*, 2011. **696**: p. 205-24.
70. Chambers, M.C., et al., *A cross-platform toolkit for mass spectrometry and proteomics*. *Nat Biotechnol*, 2012. **30**(10): p. 918-20.
71. Wishart, D.S., et al., *HMDB 5.0: the Human Metabolome Database for 2022*. *Nucleic Acids Res*, 2022. **50**(D1): p. D622-D631.
72. Bjarterot, P., et al., *Met-ID: An Open-Source Software for Comprehensive Annotation of Multiple On-Tissue Chemical Modifications in MALDI-MSI*. *Anal Chem*, 2025. **97**(16): p. 9033-9041.

73. Ellis, S.R., et al., *Automated, parallel mass spectrometry imaging and structural identification of lipids*. Nat Methods, 2018. **15**(7): p. 515-518.
74. Novak, J., A. Skriba, and V. Havlicek, *CycloBranch 2: Molecular Formula Annotations Applied to imzML Data Sets in Bimodal Fusion and LC-MS Data Files*. Anal Chem, 2020. **92**(10): p. 6844-6849.
75. Guo, G., et al., *Automated annotation and visualisation of high-resolution spatial proteomic mass spectrometry imaging data using HIT-MAP*. Nat Commun, 2021. **12**(1): p. 3241.
76. Janda, M., et al., *Determination of Abundant Metabolite Matrix Adducts Illuminates the Dark Metabolome of MALDI-Mass Spectrometry Imaging Datasets*. Anal Chem, 2021. **93**(24): p. 8399-8407.
77. Bond, N.J., et al., *massPix: an R package for annotation and interpretation of mass spectrometry imaging data for lipidomics*. Metabolomics, 2017. **13**(11): p. 128.
78. Ovchinnikova, K., et al., *OffsampleAI: artificial intelligence approach to recognize off-sample mass spectrometry images*. BMC Bioinformatics, 2020. **21**(1): p. 129.
79. Semente, L., et al., *rMSIannotation: A peak annotation tool for mass spectrometry imaging based on the analysis of isotopic intensity ratios*. Anal Chim Acta, 2021. **1171**: p. 338669.
80. Baquer, G., et al., *rMSIcleanup: an open-source tool for matrix-related peak annotation in mass spectrometry imaging and its application to silver-assisted laser desorption/ionization*. J Cheminform, 2020. **12**(1): p. 45.
81. Palmer, A., et al., *FDR-controlled metabolite annotation for high-resolution imaging mass spectrometry*. Nat Methods, 2017. **14**(1): p. 57-60.
82. Harwood, T.V., et al., *BLINK enables ultrafast tandem mass spectrometry cosine similarity scoring*. Sci Rep, 2023. **13**(1): p. 13462.
83. Onoprishvili, T., et al., *SimMS: a GPU-accelerated cosine similarity implementation for tandem mass spectrometry*. Bioinformatics, 2025. **41**(3).
84. Millikan, R.A., *The General Law of Fall of a Small Spherical Body through a Gas, and its Bearing upon the Nature of Molecular Reflection from Surfaces*. Physical Review, 1923. **22**(1): p. 1-23.
85. Gabelica, V., et al., *Recommendations for reporting ion mobility Mass Spectrometry measurements*. Mass Spectrom Rev, 2019. **38**(3): p. 291-320.
86. de Cripan, S.M., et al., *Predicting the Predicted: A Comparison of Machine Learning-Based Collision Cross-Section Prediction Models for Small Molecules*. Anal Chem, 2024. **96**(22): p. 9088-9096.
87. Plante, P.L., et al., *Predicting Ion Mobility Collision Cross-Sections Using a Deep Neural Network: DeepCCS*. Anal Chem, 2019. **91**(8): p. 5191-5199.
88. Yang, F., et al., *Collision Cross Section Prediction with Molecular Fingerprint Using Machine Learning*. Molecules, 2022. **27**(19).
89. Zhang, H., et al., *AllCCS2: Curation of Ion Mobility Collision Cross-Section Atlas for Small Molecules Using Comprehensive Molecular Representations*. Analytical Chemistry, 2023. **95**(37): p. 13913-13921.
90. Guo, R., et al., *Highly accurate and large-scale collision cross sections prediction with graph neural networks*. Commun Chem, 2023. **6**(1): p. 139.

91. Wisanpitayakorn, P., et al., *Accurate Prediction of Ion Mobility Collision Cross-Section Using Ion's Polarizability and Molecular Mass with Limited Data*. Journal of Chemical Information and Modeling, 2024. **64**(5): p. 1533-1542.
92. Engler Hart, C., et al., *Evaluating the generalizability of graph neural networks for predicting collision cross section*. J Cheminform, 2024. **16**(1): p. 105.
93. Xie, T., et al., *Large-scale prediction of collision cross-section with very deep graph convolutional network for small molecule identification*. Chemometrics and Intelligent Laboratory Systems, 2024. **252**: p. 105177.
94. Bouwmeester, R., et al., *Predicting ion mobility collision cross sections and assessing prediction variation by combining conventional and data driven modeling*. Talanta, 2024. **274**: p. 125970.
95. Karnovsky, A. and S. Li, *Pathway Analysis for Targeted and Untargeted Metabolomics*. Methods Mol Biol, 2020. **2104**: p. 387-400.
96. Jeuken, G.S. and L. Kall, *Pathway analysis through mutual information*. Bioinformatics, 2024. **40**(1).
97. Harris, C.R., et al., *Array programming with NumPy*. Nature, 2020. **585**(7825): p. 357-362.
98. Virtanen, P., et al., *SciPy 1.0: fundamental algorithms for scientific computing in Python*. Nat Methods, 2020. **17**(3): p. 261-272.
99. Matsakis, N.D. and F.S. Klock, II, *The rust language*, in *ACM SIGAda Ada Letters*. 2014. p. 103-104.
100. TensorFlow-Developers. *Tensorflow v2.20.0*. 2025; Available from: <https://www.tensorflow.org/>.
101. Peters, T. *PEP 20 - The Zen of Python*. 2004; Available from: <https://peps.python.org/pep-0020/>.
102. Perkel, J.M., *Why scientists are turning to Rust*. Nature, 2020. **588**(7836): p. 185-186.
103. Foundation, G., *GTK*. 1997, <https://www.gtk.org/>.
104. Ernerfeldt, E., *Egui: an easy-to-use GUI in pure Rust*. 2020, <https://docs.rs/egui/0.31.1/egui/>.
105. Héctor. *iced v0.14.0: A cross-platform GUI library inspired by Elm*. 2025; Available from: <https://github.com/iced-rs/iced>.
106. Atlassian. *Bitbucket*. Available from: <https://bitbucket.org>.
107. Vicari, M., et al., *Spatial multimodal analysis of transcriptomes and metabolomes in tissues*. Nat Biotechnol, 2024. **42**(7): p. 1046-1050.
108. RDKit. *Open Source Cheminformatics*. Available from: <https://www.rdkit.org>.
109. Pedregosa, F., et al., *Scikit-learn: Machine Learning in Python*. Journal of Machine Learning Research, 2011. **12**: p. 2825-2830.
110. Kluyver, T., et al., *Jupyter Notebooks -- a publishing format for reproducible computational workflows*, in *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, F. Loizides and B. Schmidt, Editors. 2016. p. 87-90.
111. Tolnay, D. *serde v1.0.228: A generic serialization/deserialization framework*. 2025; Available from: <https://github.com/serde-rs/serde>.

112. Stone, J. *rayon v1.11.0: Simple work-stealing parallelism for Rust*. 2025; Available from: <https://github.com/rayon-rs/rayon>.
113. Rusqlite-developers. *rusqlite v0.38.0: Ergonomic wrapper for SQLite*. 2025; Available from: <https://github.com/rusqlite/rusqlite>.
114. Bostock, M. *D3.js - Data-Driven Documents*. 2012; Available from: <http://d3js.org/>.
115. Probst, D. and J.L. Reymond, *SmilesDrawer: Parsing and Drawing SMILES-Encoded Molecular Structures Using Client-Side JavaScript*. J Chem Inf Model, 2018. **58**(1): p. 1-7.
116. Johnson, S.G. *The NLOpt nonlinear-optimization package*. 2007; Available from: <https://github.com/stevengj/nlopt>.
117. Powell, M.J.D., *The NEWUOA software for unconstrained optimization without derivatives*, in *Large-Scale Nonlinear Optimization*, G.D. Pillo and M. Roma, Editors. 2006, Springer. p. 255-297.
118. Powell, M.J.D., *The BOBYQA algorithm for bound constrained optimization without derivatives*. 2009, Department of Applied Mathematics and Theoretical Physics, Cambridge University: Cambridge, UK.
119. Mockers, F. *bevy 0.18: A refreshingly simple data-driven game engine and app framework*. 2025; Available from: <https://github.com/bevyengine/bevy>.
120. Zhou, Z., et al., *Ion mobility collision cross-section atlas for known and unknown metabolite annotation in untargeted metabolomics*. Nat Commun, 2020. **11**(1): p. 4334.

Generative AI

Generative AI tools such as GPT4 and 5 [1], (OpenAI) were used to generate some of the code for the projects presented in this thesis. Generative AI was not used to generate any of the text in this thesis, nor in any of the projects. I take full responsibility for the contents of this thesis.

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Pharmacy 397

Editor: The Dean of the Faculty of Pharmacy

A doctoral dissertation from the Faculty of Pharmacy, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Pharmacy. (Prior to January, 2005, the series was published under the title “Comprehensive Summaries of Uppsala Dissertations from the Faculty of Pharmacy”.)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-575210



ACTA UNIVERSITATIS
UPSALIENSIS
2026