Deep Metabolome Profiling Using mzmine and Self-Supervised Deep Learning Molecular Networking based on LC-TIMS-MS2

Ansgar Korf [1]; Robin Schmid [1]; Steffen Heuckeroth [1]; Roman Bushuiev [2]; Tomáš Pluskal [1,2]; Aiko Barsch [3]; Nikolas Kessler [3]; Heiko Neuweger [3]; Sven W. Meyer [3]; Claudia Martelli [4]; Matthew R. Lewis [3]

[1] mzio GmbH, Bremen, Germany; [2] Institute of Organic Chemistry and Biochemistry of the Czech Academy of Science, Prague, Czech Republic; [3] Bruker Daltonics GmbH & Co.KG, Bremen, Germany; [4] Bruker Switzerland AG, Fällanden, Switzerland

Introduction

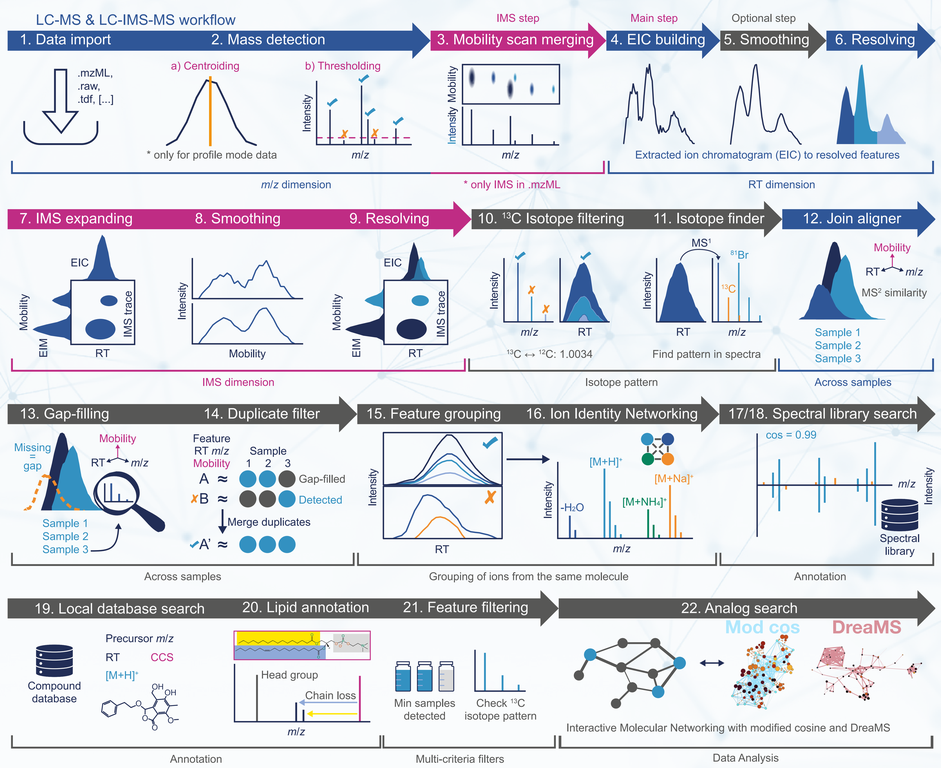
Modern metabolomics using LC-MS2 generates rich, multidimensional datasets, yet only a fraction of features can typically be annotated. The introduction of the **TIMS-MX analyzer on the timsMETABO** platform adds a new scan mode - **MoRE (Mobility Range Enhancement)** - which further increases peak capacity beyond traditional LC-MS2 and previous LC-TIMS-MS2 workflows. This enables cleaner, more informative MS2 spectra and enhances structural annotation capabilities.

We present a computational workflow to process complex multi-dimensional **LC-TIMS-MS2** datasets, acquired with **MoRE** for deep MS2 coverage. The open-source MS data processing platform **mzmine**[1] provides workflows combining spectral processing, feature detection, compound annotation, and molecular networking to map the chemical space.

Additionally, we incorporate **self-supervised deep learning** for MS2 spectral similarity using **DreaMS**[2] **(Deep Representations Empowering the Annotation of Mass Spectra)**, a transformer-based neural network. This enables the discovery of structurally related compounds, even in the absence of direct spectral matches, improving annotation depth and network connectivity.

Together, this pipeline enables deeper, faster, and more interpretable metabolome profiling, unlocking new dimensions of chemical space in untargeted metabolomics..

Data Processing



The **mzmine** data processing workflow **extracts**, **aligns**, and **annotates features** from LC-MS2 and LC-TIMS-MS2 datasets to enable comprehensive metabolomic profiling. It transforms raw instrument data into structured information by:

- **Detecting features** across *m/z*, retention time, and, when available, ion mobility dimensions
- **Aligning features** across multiple samples to enable quantitative comparisons
- **Annotating features** using isotope patterns, adduct relationships, and MS2 spectral library matches using the open librarys MSnLib and MassBank of North America (MoNA)
- Building **interactive ion identity and feature based molecular networks** to search for analog compounds

The result is a richly annotated dataset for statistical analysis, biological interpretation, and molecular networking, for both targeted and untargeted metabolomics.

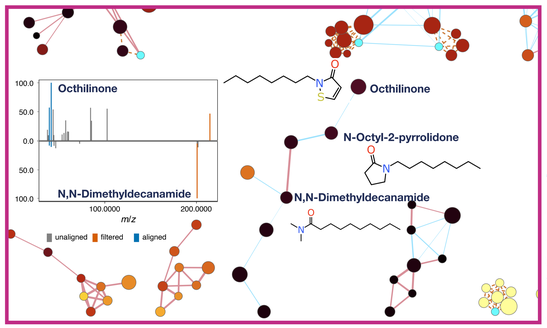
Methods

NIST SRM 3672 smokers' urine was diluted 1:10 and subsequently analyzed by **LC-MS2** and **LC-TIMS-MS2** using the **timsMETABO** instrument. For the LC-MS2 dataset, the instrument was operated as a QTOF with TIMS deactivated. MS2 spectra were acquired in **autoMSMS** acquisition mode. For the LC-TIMS-MS2, the novel scan mode **MoRE** was utilized to acquire mobility-resolved MS2 spectra.

Both datasets were processed with **mzmine 4.6.1** by utilizing the **mzwizard** for streamlinined workflow definition. These results were used for a comparison of regular QTOF performance and a TIMS-enhanced analysis.

Furthermore, the processed datasets were used to generate **interactive molecular networks** within mzmine, including **ion identity and feature-based molecular networking (FBMN). For FBMN, we scored MS2 spectral similarity based on both modified cosine** and **DreaMS** similarity**.**

The MS2 spectra of **Octhilinone** and **N,N-Dimethyldecanamide** show low similarity (modified cosine score = 0.346) and are not grouped by traditional molecular networking. **DreaMS** bridges this gap, clustering both structurally similar compounds within the same network.

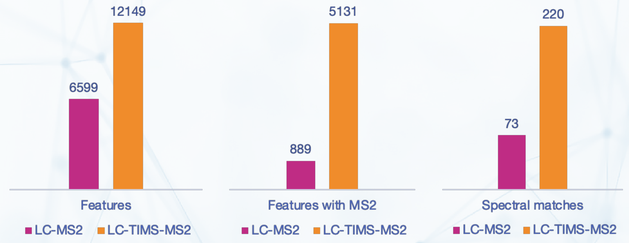


LC-MS2 and LC-TIMS-MS2 comparison

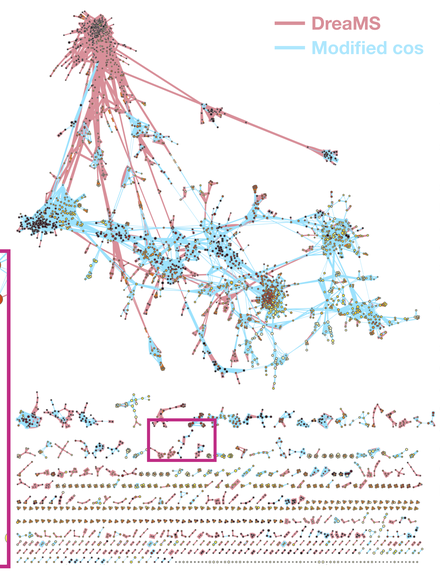
Higher feature count with LC-TIMS-MS2:
LC-TIMS-MS2 detected **~84% more features** than LC-MS2 (12149 vs. 6599), indicating enhanced separation power and sensitivity through ion mobility.

Substantial gain in MS2 coverage:
The number of features with MS2 spectra increased by over **5.7× in LC-TIMS-MS2** (5131 vs. 889), reflecting improved sampling of the ion population.

Improved spectral annotation:
Spectral matches more than **triple** in LC-TIMS-MS2 (220 vs. 73), suggesting better identification potential and enhanced annotation confidence.



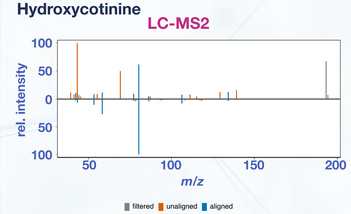
Interactive Molecular Networking



Ion identity and feature-based molecular networks can be interactively explored within mzmine. For FBMN, multiple MS2 similarity measures can be applied simultaneously. Here, **modified cosine** (blue) and **DreaMS** (red) were **combined** to maximize structural insights. Modified cosine effectively groups molecules differing by a single modification, while deep learning approaches like **DreaMS** enable **clustering of structurally related compounds with distinct fragmentation patterns**.

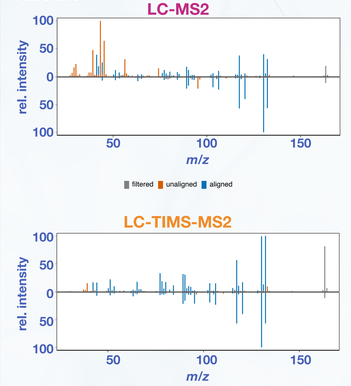
MS2 spectra comparison

Hydroxycotinine



The **LC-MS2 spectrum** of Hydroxycotinine contains **multiple unaligned signals (orange)** compared to the library spectrum, leading to a low similarity score and a potentially missed annotation. In contrast, the mobility resolved **LC-TIMS-MS2 spectrum** shows **minimal misalignment**, yielding a significantly higher similarity score.

Nicotine



A similar pattern is observed for Nicotine. The LC-MS2 spectrum contains unaligned signals relative to the LC-TIMS-MS2 spectrum. Moreover, the **LC-TIMS-MS2 spectrum** displays **enhanced signal intensity in the low *m/z* range compared to the matched library** spectrum.

Conclusion

Two metabolomics LC-MS profiling methods were developed on the **timsMETABO** instrument, one with **TIMS** using the novel **MoRE** MS2 scan option, and the other one without TIMS, using conventional DDA MS2 acquisition. Both methods were used to analyze NIST SRM 3672 smokers' urine. The data was subsequently processed and compared with **mzmine 4.6.1**.
Using TIMS and the MoRE scan mode resulted in higher peak capacity, reflected by **~84% more detected features, over 5.7 x more features with corresponding MS2 spectra, and in 3 x higher number of spectral matches**, showing the advantage of TIMS-based methods, compared to conventional LC-QTOF methods. Furthermore, mobility resolved MS2 spectra yield higher similarity scores and less chimeric MS2 spectra. Molecular networking revealed more complete and structurally coherent clusters with TIMS-enhanced data, and the integration of **DreaMS** further improved clustering of structurally related compounds with divergent fragmentation patterns.

[1] Schmid, R., Heuckeroth, S., Korf, A. *et al*. Integrative analysis of multimodal mass spectrometry data in MZmine 3. Nature Biotechnology 41, 447–449 (2023)
[2] Bushuiev R *et al*., Self-supervised learning of molecular representations from millions of tandem mass spectra using DreaMS. Nature Biotechnology 2025, doi: 10.1038/s41587-025-02663-3

Conflict of interest statement: A.K., S.H., and R.S. are co-founders and employees at mzio GmbH. T.P. is a co-founder of mzio GmbH. mzio GmbH develops analytical software, such as mzmine, which was used in this work. The development of mzmine is also funded by the European Union, the BAB - Funding Bank for Bremen and Bremerhaven, and the Senator of Economics, Ports and Transformation Bremen (65002459). N.K., A.B., H.N., C.M., S.W.M., and M.R.L. are employees of Bruker Corporation or one of its subsidiaries ("Bruker"). Bruker manufactures and sells analytical instruments including mass spectrometers and software. Bruker mass spectrometers and software were used in this study.