**PhD Position: Developing Next-Generation LC-MS/MS and Mass Spectrometry Imaging Data Processing Pipelines**

Mass spectrometry has revolutionized molecular exploration of life events by enabling the profiling of proteins, glycoproteins, lipids, drugs, and metabolites in biological samples—information that is often inaccessible or only partially predictable from genomic or transcriptomic data. These molecules are chemically far more complex than DNA or RNA base pairs, making their analysis highly challenging.

Mass spectrometry has become a central analytical platform for fundamental and clinical research, capable of delivering deep molecular profiles, including tens of thousands of lipids, metabolites, and proteins from each clinical sample. The large and highly complex datasets generated by modern mass spectrometry platforms—whether coupled with liquid chromatography (LC-MS/MS) or applied in imaging mode (mass spectrometry imaging, MSI)—contain vast amounts of compound-related information, much of which remains unexplored.

Over the past decades, we have developed the **PASTAQ** library and toolset, providing a generic framework for the quantitative processing of metabolomics, lipidomics, glycomics, and proteomics LC-MS/MS data acquired with data-dependent acquisition (DDA), as well as **MALDIViewer**, a tool for the analysis of complex MSI datasets. Both tools are developed in C++ with Python bindings.

**PASTAQ Overview:**

**Pipelines and Systems for Threshold-Avoiding Quantification of LC-MS/MS Data (PASTAQ)** is a platform designed for the quantitative preprocessing, exploration, and annotation of LC-MS/MS data, with a special emphasis on full data traceability and the quantification of low-intensity signals. LC-MS/MS data can be conceptualized as a 2D image, where compounds are separated by their chemical properties (via liquid chromatography) and mass (via mass spectrometry), appearing as 2D Gaussian-like peaks. The pipeline’s task is to detect and quantify these peaks.

**MALDIViewer Overview:**

MALDIViewer is based on a highly efficient data structure built from indexed triplets (m/z, pixel index, intensity) of MSI datasets, coupled with algorithms that enable accurate, data-driven ion image extraction limited only by mass resolution. The tool also incorporates **MSIWarp**, a spectral alignment algorithm that automatically corrects mass shifts across billions of acquired spectra.

**PhD Research Description:**

The PhD project involves the development of new algorithms and the extension of functionality for these data processing pipelines.

**PASTAQ LC-MS/MS Development:**

* Develop efficient peak detection algorithms capable of handling asymmetry and distortions in Gaussian peak shapes, as well as fast methods for detecting and deconvolving overlapping peak clusters.
* Create novel algorithms for data-independent acquisition (DIA), enabling efficient extraction of pseudo-spectra from highly complex and mixed LC-MS/MS datasets using deep learning approaches.
* Develop methods to annotate PASTAQ quantified peaks with untargeted metabolite identifications.
* Adapt the LC-MS/MS data processing and visualization pipeline for GPU acceleration using CUDA, OpenGL compute, or Vulkan API.

**MALDIViewer MSI Data Processing:**

* Integrate (image registration) MSI data with high-resolution microscopy images to transfer cellular and tissue features detected across both modalities.
* Develop efficient AI-based feature detection methods using large language models (LLMs) and convolutional neural networks (CNNs) on combined MSI and microscopy datasets.
* Annotate MSI datasets efficiently with metabolite and lipid identifications obtained from LC-MS/MS data.
* Accelerate the computation and visualization of MSI datasets using GPU technologies (CUDA, OpenGL compute, Vulkan API).

**Project Expertise Level:**

More information is available at:

* [PASTAQ general overview](https://pastaq.horvatovichlab.com/)
* [PASTAQ GitHub repository](https://github.com/PASTAQ-MS/PASTAQ)
* [MSIWarp GitHub repository](https://github.com/horvatovichlab/MSIWarp)

**Supervision and Contact:**

The project will be supervised by:

* Dr. Peter Horvatovich (p.l.horvatovich@rug.nl), Analytical Biochemistry, GRIP, University of Groningen (RUG)
* Dr. Venus Soancatl Aguilar (v.soancatl.aguilar@rug.nl), Data Analysis Team, CIT, University of Groningen (RUG)

**Required Expertise:**

* Strong background in C++, Python, and/or R programming
* Experience in signal processing
* Familiarity with deep learning and large language model (LLM) packages
* Background in physics, computer science, or a related field