Mass spectrometry is a laboratory technique that allows to measure the mass to charge ratio of chemical compounds. Intuitively, we can assume that it allows for precise weighting of individual molecules.

Mass spectrometry has become a fundamental tool in proteomics, metabolomics, lipidomics, and other high-throughput studies of complex biochemical samples. The possibility of investigating complex samples like tissues, foods and body fluids is crucial in both fundamental research and applied biomedical sciences. This important high-throughput technology stimulates the development of entire branches of computational sciences such as proteomics, lipidomics, and computational metabolomics, whose task is to develop algorithms, formal models and statistical methods for analyzing spectrometric data.

It should be emphasised that the development of high-throughput experimental techniques in biomedical science provides huge amounts of data, that are very difficult for interpretation. Interdisciplinary teams of chemists, biologists, physicians, mathematicians and IT professionals are required to perform the effective analysis of such datasets. Only the unique combination of the broad chemical knowledge, the mathematical culture and algorithmic skills can provide efficient solutions accepted by the scientific community.

In this project we focus on data obtained in mass spectrometers. The research tasks planned in the project concern the analysis of high resolution spectra, as recent advances in instrumental design clearly show the tendency to construct much more affordable mass spectrometers with still higher resolution

Computational methods in mass spectrometry have to keep up with the technological progress. It is often necessary to propose new mathematical and algorithmic solutions. High resolution mass spectrometry is a field where effective algorithms are lacking, and we plan to contribute to the field at all stages of data processing: from innovative idea of using Wasserstein distance to compare spectra, through algorithms for analyzing overlapping signals, identifying the atomic composition of the detected substances, to high-level metabolic pathway analysis.

We hope that the solutions proposed by us will be specially useful for processing data from mass spectrometric imaging. This is a new technique used in medical diagnostics where the entire tissue is examined. Moreover we believe that the automatic identification of the atomic composition will allow for automatic annotation of metabolites contained in the test sample. The most challenging task we want to take is an attempt to reconstruct metabolic pathways based on spectrometric data.