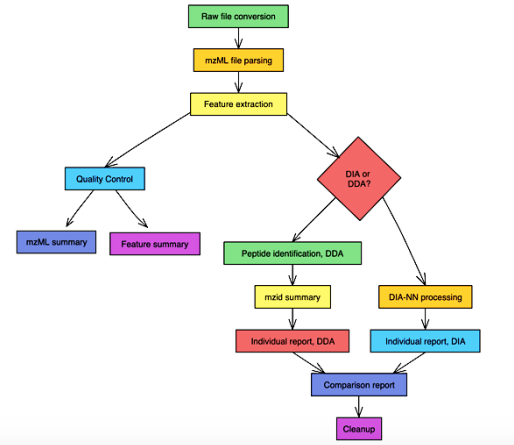
**Saghar Toresson**

# Optimal Settings for Maximal Proteome Coverage in Mass Spectrometry

In the field of biological sciences, the study of proteins, known as proteomics, is essential for advancing our understanding of health and disease. One of the key tools in this field is the mass spectrometer, an instrument that can precisely analyze the types and amounts of proteins in a sample. Proteins are vital to every cell and are involved in nearly all biological processes, from speeding up chemical reactions to fighting infections.

There are two techniques in mass spectrometer: Data-Dependent Acquisition (DDA) and Data-Independent Acquisition (DIA). DDA targets and analyzes the most abundant proteins in a sample, which makes it great for looking at well-known proteins but can miss less common ones. DIA, on the other hand, captures information on all proteins present, not just the most abundant, providing a fuller picture of what's in the sample.

To manage and interpret the large amount of data, I developed a computational pipeline using a program called Snakemake in my project. This pipeline automates the entire process: from converting the raw data, extracting important features, to running quality control and identifying peptides, building blocks of proteins. This automation significantly speeds up the research process, allowing for quicker and more accurate results. The workflow is adept at managing both Data-Dependent Acquisition (DDA) and Data-Independent Acquisition (DIA) methods, allowing it to adapt the analysis based on the data type.




By adjusting how the mass spectrometer is set up, we can detect a wider range of proteins, including those that are less common. We also could compare which settings worked better for our data. Our tests have shown that using narrower and overlapping windows during DIA analysis improves protein detection and this can be applied to similar datasets. This developed method is invaluable for future research by helping us better understand the complex world of cellular functions.

Reserchers in academic and industrial research focusing on proteomics can use this pipeline to enhance the precision and efficiency of their protein analyses.

Master's Degree Project in Bioinformatics 45 credits 2024
Deaprtment of immunolotechnology, Lund University

Advisor: **Fredrik Levander**