Summary

Proteins are the building bricks of cells, tissues and organs in all living organisms. Depending on which proteins are expressed in a cell and to what extent, defines the function of the cell and the eventual tissue. Not only expression but post-translational modifications of the proteins as well, will define the function of the proteins. Through proteomics, the research field that analysis the proteins and their interactions within a given sample, insights in the mechanisms present in a cell are obtained. Because of its untargeted analysis, LC-MS/MS is a highly appreciated proteomics analysis tool capable of identifying the proteins present in a sample next to detecting differences in protein abundancies. The consequence of its popularity is the emerging need for throughput next to the ever urging need of pushing the detection limits. Although a lot of improvements have been applied already in the past decades and the instrumentation has evolved at revolutionary speed, there is still a need for higher throughput without compromising quality.

There are several steps in an LC-MS/MS proteomics analysis in which adjustments can be made to increase the throughput. An LC-MS/MS proteomics analysis consists of a sample preparation step, an LC-MS/MS step and a data analysis step. Here we focused on the improvement of throughput on the LC step and the consequences on the quality of the data in using the higher throughput data-independent acquisition (DIA) method on the MS step.

On the LC side a dual column setup was implemented with the comparison of the different options next to comparison with the original more default setup. The dual column setup is capable of doubling the duty cycle without compromising data quality. Moreover the dual column setup even improves the analysis quality by decreasing sample-to-sample column memory. Its superiority towards a single column setup compensates for the higher complexity it implies. Indeed the setup is less straight forward and the more the duty cycle needs to be increased, the more complex the setup becomes, eventually being incompatible with daily operations. Therefore we limited the improvements to the point it is still manageable in a day-to-day operation mode, not putting the duty cycle to the maximum though.

On the MS side, an emerging data acquisition mode, data-independent acquisition (DIA) opens up opportunities towards higher throughput and improved sensitivity. The implementation of DIA was stalled for a long time because of the lack of descent data analysis options. In recent years different approaches emerged to analyze DIA data. Although the data resides from the same run, the different ways of analyzing the data comes with different outcomes. Our concern about the different outcomes was the quality of the different results. Therefore we benchmarked different methods of data analysis of DIA data through our own benchmark sample on the level of sensitivity, reproducibility and accuracy. A focus was put on a quite recent way of analyzing DIA data through the use of *in silico* predicted libraries. The use of these *in silico* predicted libraries renders high sensitivity, reproducibility with an acceptable accuracy while providing an independent way of creating the spectral library. This liberty paves the way of DIA analysis of unconventional prepared samples, in which different proteases are used, different chemical reactions are applied, and many more.

In summary, combining the high-throughput dual column LC system with DIA analysis using *in silico* predicted spectral libraries shifts LC-MS/MS based proteomics to the fast lane without compromising sensitivity, reproducibility or accuracy.

About the author

An Staes started her scientific career in the early 2000's in the Gevaert Lab in the field of N-terminomics. The development and update of the N-terminal enrichment protocol used in the lab (N-terminal COFRADIC) was her main objective while digging deeper into the world of mass spectrometry driven proteomics. Following the path of the mass spectrometer evolution she became an expert in LC-MS proteomics. She serves the field with her expertise by taking the lead in developing and applying the rapidly evolving methods in the VIB Proteomics Core facility that arose from the Gevaert lab.

Through her recent role as president of the Belgian proteomics association (BePA) she embraces pushing the Belgian proteomics society forward. She decided to give her expertise that extra push by obtaining a PhD in the field of mass spectrometry driven proteomics under the guidance of Prof. Dr. Kris Gevaert, Prof. Dr. Francis Impens and Dr. Simon Devos. In her PhD she dug deeper into the world of high throughput MS without compromising quality by implementing a dual column LC system and assessing DIA data analysis workflows.

GHENT UNIVERSITY

FACULTY OF MEDICINE AND HEALTH SCIENCES

VIB-UGENT CENTER FOR MEDICAL BIOTECHNOLOGY

VIB PROTEOMICS CORE

Contact

Department of Biomolecular Medicine, Ghent University
VIB-UGent Center for Medical Biotechnology, Ghent
VIB Proteomics Core, Ghent, Belgium

an.staes@ugent.be
T +32 9 224 9933
www.ugent.be

Promotors

Prof. Dr. Kris Gevaert
Department of Biomolecular Medicine, Ghent University
VIB-UGent Center for Medical Biotechnology, Ghent

Prof. Dr. Francis Impens
Department of Biomolecular Medicine, Ghent University
VIB-UGent Center for Medical Biotechnology, Ghent

Dr. Simon Devos
Department of Biomolecular Medicine, Ghent University
VIB-UGent Center for Medical Biotechnology, Ghent
VIB Proteomics Core, Ghent

Members of the examination committee

Prof. Dr. Katleen De Preter
Department of Biomolecular Medicine, Ghent University (Chair)

Dr. Daria Fijalkowska
Department of Biomolecular Medicine, Ghent University

Dr. Apr.Sigrid Verhelst
Department of Pharmaceutics, Ghent University

Prof.Dr.Lynn Vanhaecke
Department of Translational Physiology, Infectiology and Public Health, Ghent University (Secretary)

Prof. Dr. Liesbet Temmerman
Department of Biology, KU Leuven

Prof.Dr. Claire Eyers
Department of Biochemistry and Systems Biology, ISMIB, University of Liverpool

Shifting mass spectrometry based proteomics to the fast lane

An interplay between high throughput LC and data independent acquisition without compromising depth, precision or accuracy

An Staes

Dissertation submitted to obtain the degree
'Doctor in Health Sciences'
Academic year 2024-2025