Automatic Peptides Selection for Targeted Proteomics

Yassene Mohammed[1,2], Dominik Domański[3], Angela M. Jackson[1], Derek S. Smith[1], André M. Deelder[2], Magnus N. Palmblad[2], and Christoph H. Borchers[1]

[1]University of Victoria - Genome British Columbia Proteomics Centre,
University of Victoria, Canada
[2]Center for Proteomics and Metabolomics,
Leiden University Medical Center, the Netherlands
[3]Mass Spectrometry Laboratory,
Institute of Biochemistry and Biophysics PAS, Warsaw, Poland

Abstract. Multiple Reaction Monitoring (MRM), also known as Selected Reaction Monitoring (SRM), is a powerful targeted proteomics method. The approach is based on the generation of specific quantitative assays for each protein of interest and can be used to accurately quantitate large sets of proteins at high throughput. One challenge in Multiple Reaction Monitoring Proteomics analysis is however to select the most appropriate surrogate peptides to represent a target protein. Several features are required for a good surrogate peptide -- it should be unique within the target proteome, efficiently liberated during enzymatic digestion, and free of post translational modifications.

We use established eScience practice to develop a software package to automatically generate the most appropriate surrogate peptides to represent target proteins in an LC/MRM-MS analysis. Our method integrates information from different online repositories about the proteins, their tryptic peptides and suitability for MRM. The software is built using scientific workflows engine, i.e. Taverna, it connects to multiple online repositories, from which the information about proteins and peptides are retrieved and integrated. In comparison with manual selection of peptides, our method enriched the selection, eliminated human error and allowed faster generation of peptide selection lists.

Because the software was developed using scientific workflow, it allows data provenance, reproducibility of calculation and experiment, as well as sharing the analysis method with other scientists. All are properties on which the e-Science community worked for few years now to make them ready for use in science.

Keywords: *MRM, SRM,* targeted proteomics, peptide selection, data integration, scientific workflow