Protein Prospector as a Component in a Label-Free/SRM Pipeline

Peter R. Baker[1], Anatoly Urisman[1] and Robert J. Chalkley[1]

[1]Mass Spectrometry Facility, Dept. of Pharmaceutical Chemistry, University of California, San Francisco, USA

Introduction

Construction of reference spectral libraries is a major component of most shotgun label-free and targeted proteomic pipelines. The spectral libraries typically contain spectra acquired during preliminary discovery experiments or as part of quantitative experimental runs. In either case, the process involves compiling spectra from database searches and selecting the best representative spectrum for each identified component. Skyline, a popular software package for quantitative data-dependent and targeted applications, relies on the use of reference spectral libraries formatted by the BiblioSpec program (.blib format). Previously, generation of blib-formatted libraries from Protein Prospector output required time- and storage-consuming file format conversion into mzXML and pepXML followed by BiblioSpec formatting. The blib format is now an output option in Protein Prospector. In addition the spectral viewer program MS-Viewer[1] has now been extended to display BiblioSpec and NIST libraries along with maxQuant results. MS-Viewer can also support peak list formats such as mgf, mzData, pkl, dta, mzML, mzXML, ms2 and apl and results file formats such as Protein Prospector (including crosslinking results), X!Tandem, Mascot CSV, pepXML, Thermo MSF and PRIDEXML.

Methods

Protein Prospector's Search Compare can now create output files in the BiblioSpec blib format. These files are compatible with Skyline enabling Protein Prospector to be part of a data processing pipeline for label-free and targeted quantitation. Additionally, Protein Prospector formatted blib libraries can be used for spectral library searches using BiblioSpec. Search Compare can combine results from multiple data sets and select the best spectrum for any given peptide based on p-value and many advanced filtering options. Retention time normalization can increase the compatibility of discovery spectra from multiple experiments. Full automation can be achieved via command line options.

MS-Viewer is a program within the Protein Prospector package, which has recently been extended to support upload and viewing of blib-formatted libraries in addition to NIST msp libraries and maxQuant datasets. A number of useful filtering options are now available in MS-Viewer to remove unwanted columns and rows.

Conclusion

Protein Prospector including MS-Viewer and the database search programs required to build spectral libraries is available at http://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msviewer or for local installation by emailing ppadmin@cgl.ucsf.edu.

Example MS-Viewer datasets including spectral libraries are at http://prospector.ucsf.edu/prospector/html/misc/viewereg.htm. A MaxQuant dataset will be added shortly.

A video explaining MS-Viewer is available at http://vimeo.com/30462677.

References

1. Baker PR and Chalkley RJ, MS-Viewer: A Web Based Spectral Viewer for Proteomics Results. Molecular and Cellular Proteomics, doi:10.1074/mcp.O113.037200 (2014)

Acknowledgements

MS-Viewer Display of Libraries and maxQuant Results

Columns can be removed from the displayed report.

Rows can be removed based on cell contents.

Spectral viewer parameters.

Reports can be sorted based on the contents of up to 4 columns.

Default parameters For MS-Tag search.

MS-Viewer Display and Parameter Settings showing row and column filtering options

MS-Viewer displaying BiblioSpec BSA library created from selecting the best spectra from 26 data sets using Search Compare

MS-Viewer displaying the NIST BSA library filtered to show only tryptic peptides

MS-Viewer displaying maxQuant results with some of the 61 original columns removed. The spectral display is showing a comparison of different site localization possibilities

RT Normalization

Retention time normalization of peptides between runs is very important for shotgun label-free and targeted proteomics. Currently the retention times of the earliest and latest eluting peptides are used if retention time normalization is required. A minimum of 50 peptides must be detected before normalization can take place to account for problem runs with only a few peptide hits.

Whilst there is definitely room for improvement, e.g. gradient normalization based on previous known IDs, even with this simple linear approach we are able to achieve excellent agreement between very diverse libraries enabling the use of multiple libraries simultaneously when scheduling new targeted runs. In the example shown on the right we were able to use an initial window of +/- 15 units (0-100 scale), scale it appropriately to a new 4 hr gradient and achieve >90% detection of ~150 targets (derived from both libraries).