MASPECTRAS: a platform for management and analysis of proteomics LC-MS/MS data

Hartler J[1,2], Ubaida Mohien C[1], Thallinger GG[1], Stocker G[1], Sturn A[1], Burkard TR[1,3], Körner E[4], Rader R[1], Schmidt A[5], Colinge J[6], Mechtler K[3] & Trajanoski Z[1]

[1] Institute for Genomics and Bioinformatics, Graz University of Technology, Graz, Austria
[2] Austrian Research Centers GmbH –ARC, eHealth Systems, Graz, Austria
[3] Research Institute for Molecular Pathology, Vienna, Austria
[4] FH Joanneum, Kapfenberg , Kapfenberg, Austria
[5] Christian-Doppler Laboratory for Proteome Analysis, Vienna, Austria
[6] Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria

Abstract:

The advancements of proteomics technologies have led to a rapid increase in the number, size and rate at which datasets are generated. Managing and extracting valuable information from such datasets requires the use of data management platforms and computational approaches.
We have developed the MAss SPECTRometry Analysis System (MASPECTRAS) [1], a platform for management and analysis of proteomics LC-MS/MS data. It is a web-based database application fulfilling the Minimum Information About a Proteomics Experiment (MIAPE) standard [2]. The platform is scalable and enables the distribution of computationally intensive tasks to a computing cluster. MASPECTRAS analysis modules include (Fig. 1): 1) import and parsing of the results from several search engines; 2) peptide validation, 3) grouping of proteins; and 4) relative quantification. Unique features of the system are the merging of results originating from different search engines, a chromatogram viewer and export to the PRoteomics IDEntifications public repository (PRIDE) [3] via PRIDE-XML. The platform is freely available at http://genome.tugraz.at/maspectras.

Methods and Results:

MASPECTRAS is based on a 3-tier architecture using the Java 2 Enterprise Edition (J2EE). A relational database (Oracle, PostgreSQL or MySQL) builds the data- or Enterprise Information System tier. The business logic is hosted by the application server JBoss. The system is accessible by web browser and it is embedded in a multi-user environment providing controlled user access to the data. It is a comprehensive, web-based data repository which has been developed on the basis of the Proteome Experimental Data Repository (PEDRo) [4] relational database and has been adapted to fulfill the MIAPE standard provided by the Proteomics Standards Initiative (PSI) [2] .

It allows the storage and comparison of results of different search engines (Sequest, Mascot, SpectrumMill, X! Tandem and OMSSA) in a single view (Fig. 2).

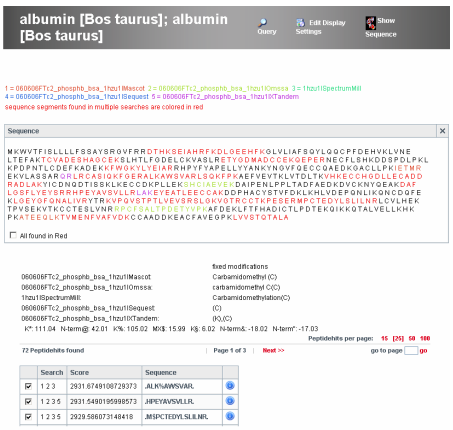

Figure 2: Combined view of the results from 5 search engines (Sequest, Mascot, SpectrumMill, X! Tandem and OMSSA). The colored top line displays the name of the search results. Peptides identified only in one of the searches have the corresponding color whereas peptides identified by multiple searches are colored red. In the last table the found peptides are listed.

A sophisticated query system allows data filtering as well as comparison of data from different searches. Validation of the data is achieved by the integration of the Peptide Prophet scoring algorithm [5] and a customizable spectrum viewer for manual verification (Fig. 3).
The protein grouping is based on Markov clustering [6] and multiple alignments [7].
For the relative quantification of the data the Automated Statistical Analysis of Protein Abundance Ratios algorithm (ASAPRatio) [8] has been integrated. The peptides are quantified automatically when the data is uploaded. The results can be checked and updated with the integrated chromatogram viewer (Fig. 4).

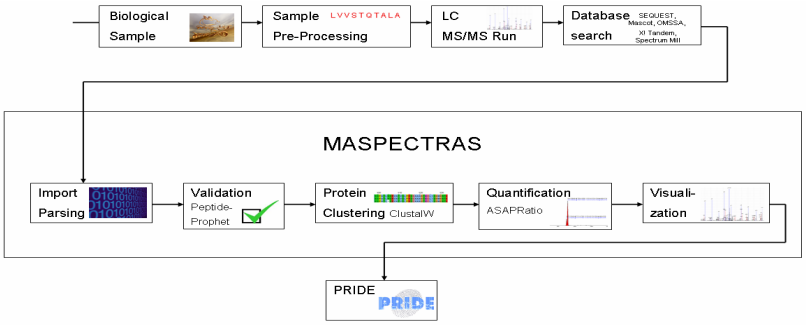

Figure 1: MASPECTRAS analysis pipeline. Search results from Sequest, Mascot, Spectrum Mill, X! Tandem, and OMSSA are imported and parsed. In the next steps peptides are validated using PeptideProphet [5] and the corresponding proteins are clustered based on Markov clustering using BLAST[6], the sequences are aligned with CLUSTAL W [7]. The peptides are quantified using the ASAPRatio algorithm [8], results are stored in the database and exported to the public repository PRIDE [3].

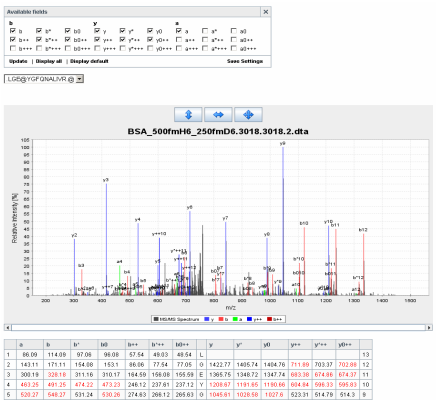

Figure 3: The spectrum viewer offers the selection of the displayable ion series, selection of the peptide hit to display, zooming- and printing possibilities.

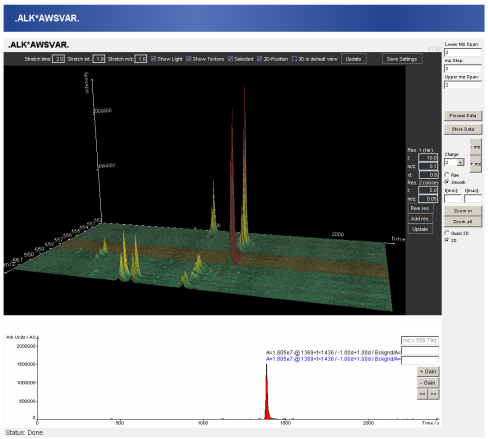

Figure 4: Chromatogram viewer for the relative quantification. The first view offers a 3D view on the raw data used for the chromatogram calculation, the second one shows the chromatogram in 2D. The range used for the 2D view is colored in dark yellow. The region used for the quantification is displayed in red. Additional peaks (regions) can be added, and stored ones can be removed manually. The chromatogram viewer allows changing the m/z step-size for a chromatogram, m/z range to be displayed next to the chromatogram of interest and the charge state, and has several zooming abilities.

Discussion:

MASPECTRAS has been evaluated using large-scale proteomics data and quantitative data from a controlled experiment [1]. In these studies the platform proved to be an efficient environment for the rapid analysis of large-scale proteomics experiments in a MIAPE compliant manner. It features an automated analysis pipeline and unique analysis tools, and provides an easy export functionality for the submission of data to public repositories. Due to its modular design and flexibility, the future requirements in proteomics data management and analysis can be easily accommodated.

References:

[1] Hartler J, et al.: MASPECTRAS: a platform for management and analysis of proteomics LC-MS/MS data. BMC Bioinformatics 2007; 8:197.
[2] Orchard S, et al.: The proteomics standards initiative. Proteomics 2003; 3(7):1374-1376.
[3] Martens L, et al.: PRIDE: the proteomics identification database. Proteomics 2005; 5(13):3537-3545.
[4] Taylor CF, et al.: A systematic approach to modeling, capturing and disseminating proteomics experimental data. Nature 2003:21, 247-254.
[5] Keller A, et al.: Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search, Anal Chem 2002; 74: 5383-5392.
[6] Enright AJ, et al.: An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res 2002; 30(7):1575-1584.
[7] D. Higgins, et al.: Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994; 22: 4673-4680.
[8] Li XJ, et al.: Automated Statistical Analysis of Protein Abundance Ratios from Data Generated by Stable-Isotope Dilution and Tandem Mass Spectrometry. Anal. Chem 2003; 75: 6648-6657.

Acknowledgements:

This work is supported by the Austrian Academy of Sciences (OEAW), the Bioinformatics Integration Network II (BIN) of the Austrian Genome Research Program (GEN-AU), the Christian-Doppler Laboratory for Genomics and Bioinformatics, and the Austrian Proteomics Platform II (APP).

Correspondence:

E-mail: juergen.hartler@tugraz.at, zlatko.trajanoski@tugraz.at
URL: http://genome.tugraz.at/maspectras