Abstract

Mass spectrometry has emerged to become a state-of-the-art methodology for the analysis of the proteome. In order to manage and extract valuable information from mass spectrometry data, a computational data management platform is indispensable. Furthermore the standardized presentation of data is of the utmost importance. At present however, no proteomics bioinformatics platform exists that fulfills the Minimum Information About a Proteomics Experiment (MIAPE) standard and that provides all the features needed for the effective analysis of mass spectrometry experiments.

In view of this we have developed a versatile MIAPE compliant platform for the management and analysis of proteomics mass spectrometry experiments: the MAss SPECTRometry Analysis System (MASPECTRAS). This platform is web-based with a database back-end and relies on the Java 2 Enterprise Edition development platform. The platform is scalable and enables the outsourcing of computationally intensive tasks to a computing cluster. The data model captures data concerning experimental design and at all other subsequent steps leading up to evaluation and result export. The analysis process relies on the output of the major mass spectrometry search engines SEQUEST, Mascot, SpectrumMill, X!Tandem and OMSSA. The data is run through an automated analysis pipeline in the following four steps: 1) import and parsing to a MIAPE compliant representation; 2) validation; 3) protein clustering; 4) peptide quantification. For manual analysis MASPECTRAS provides unique features that include the merging of results originating from different search engines, a chromatogram viewer and export to PRIDE-XML format which can be uploaded directly in a public repository. The functionality of the whole system is furthermore embedded in a multi-user environment providing controlled user access. The platform has been evaluated using large-scale and quantitative data.

The MASPECTRAS platform offers researchers an environment for the rapid analysis of large-scale proteomics experiments. Due to its modular design it is flexible enough to easily accommodate future changes in proteomics data management.

Keywords: proteomics, management platform, tandem mass spectrometry (MS/MS), MIAPE, large-scale analysis, PRIDE, AndroMDA, J2EE