Navigation, visual exploration and curation of large-scale proteomics data with Proline

David Bouyssié1*, Anne-Marie Hesse2*, Emmanuelle Mouton-Barbosoa*, Magali Rompais3*, Christine Carapito3*, Véronique Dupierriis2*, Alexandre Burel3*, Aymen Romdhani3* and Christophe Bruley2*

1 Institut de Pharmacologie et de Biologie Structurale, CNRS UMR5089, Université de Toulouse
2 Laboratoire de Spectrométrie de Masse BioOrganique (LSMBO), IPHC-OSA, Université de Strasbourg. CNRS UMR7178
* ProFi, Proteomics French Infrastructure

Due to the intrinsic complexity of bottom-up proteomics experiments, inaccuracies and errors can occur throughout the data-processing pipeline. Accordingly, result reliability must be carefully assessed, not only by statistically controlled procedures, but also through examination of the underlying data by experts. Proline software suite provides a unique environment combining a structured representation of data and metadata based on an underlying relational database with an interactive graphical user interface from which users can navigate into considerable amounts of data, allowing the examination and manual curation of all results by human experts.

**Workflow Steps**

**Import & Organize search results**: Proline can import results from Mascot, OMSSA or X!Tandem and search results and identification summaries can be merged to build a parent dataset, taking into account peptides identified in all merged datasets.

**Validate identifications**: Validation can be performed at peptide-spectrum match (PSM) and protein levels: a set of predefined filters can be applied to accept or reject a PSM or a protein depending on user-specified threshold values applied to different properties. A target-decoy validation approach can be performed by adjusting the false discovery rate (FDR) to a user value at both levels.

**LC-MS feature extraction**: Proline first detects chromatographic peaks from raw MS data contained in mzDB files and assigns those LC-MS peaks to the validated PSMs.

**Cross Assignment**: The software then aligns the LC-MS maps representing identified signals to assign unidentified peaks to peptide ions across all runs.

**Protein Quantitation**: The ion abundances are stored in the database, and Proline can then reuse the quantitative data to summarize peptide ion measurements as protein abundances using various summarization methods. A change of method does not require the whole quantification process to be restarted.

Performances have been assessed with a proteomic standard dataset composed of an equimolar mixture of 48 human proteins (UP91, Sigma) spiked at different concentrations (from 10amol to 50fmol) into a yeast cell lysate background.

**Data exploration & Visualization**

In Proline, results from each step of the workflow can be displayed and inspected, starting with the MS/MS spectra and their sequence interpretations. Validated and rejected PSMs can both be displayed and searched using different criteria.

Users can browse the content of identification results through a set of predefined views in Proline. The view represented here shows validated protein sets for an identification dataset.

The alignment process is a critical step in label-free quantification. Proline provides an effective tool to check the quality of alignments by viewing the estimated retention time difference between runs.

In the case of LC-MS quantification, peptides measured across aligned MS runs are represented together. For each peptide, the quantified ions and their extracted elution peaks can be viewed, as well as the isotope elution peaks. Suspect profiles can be individually discarded by the user, and protein quantification can then be recomputed without the invalidated measurements (as exemplified by co-eluting peaks, in the above snapshot).

Proline also offers a web graphical user interface to remotely access Proline Server via a web browser.

Navigation within data is facilitated by allowing users to define their own navigation path through the data. A navigation path is materialized by a new window layout where views are dynamically updated depending on the user’s selection.

http://proline.profiproteomics.fr/