

Sequence analysis

# VisioProt-MS: interactive 2D maps from intact protein mass spectrometry

Marie Locard-Paulet \*, Julien Parra, Renaud Albigot, Emmanuelle Mouton-Barbosa, Laurent Bardi, Odile Burlet-Schiltz and Julien Marcoux

Institut de Pharmacologie et de Biologie Structurale, Université de Toulouse, CNRS, UPS, 31077 Toulouse, France

\*To whom correspondence should be addressed.

Associate Editor: Alfonso Valencia

Received on December 21, 2017; revised on July 13, 2018; editorial decision on July 30, 2018; accepted on August 6, 2018

## Abstract

**Summary:** VisioProt-MS is designed to summarize and analyze intact protein and top-down proteomics data. It plots the molecular weights of eluting proteins as a function of their retention time, thereby allowing inspection of runs from liquid chromatography coupled to mass spectrometry (LC-MS). It also overlays MS/MS identification results. VisioProt-MS is compatible with outputs from many different top-down dedicated software. To our knowledge, this is the only open source standalone application that allows the dynamic comparison of several MS files, a prerequisite for comparative analysis of different biological conditions. With its dynamic rendering, this user-friendly web application facilitates inspection, comparison and export of publication quality 2D maps from deconvoluted LC-MS run(s) and top-down proteomics data.

**Availability and implementation:** The Shiny-based web application VisioProt-MS is suitable for non-R users. It can be found at <https://masstools.ipbs.fr/mstools/visioprot-ms/> and the corresponding scripts are downloadable at <https://github.com/mlocardpaulet/VisioProt-MS>. It is governed by the CeCILL license (<http://www.cecill.info>).

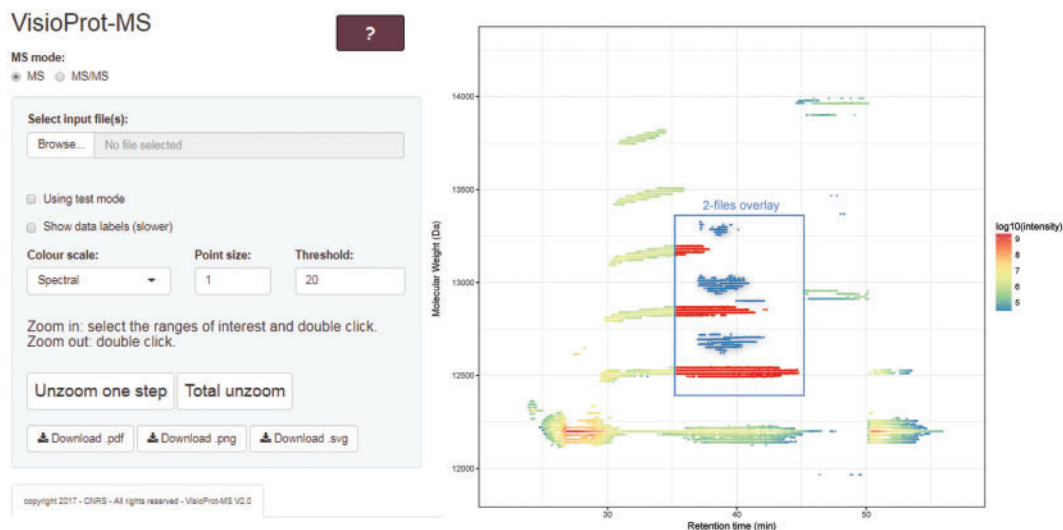
**Contact:** marie.locard@ipbs.fr

## 1 Introduction

Top-down proteomics consists in the analysis of intact proteins using liquid chromatography coupled to mass spectrometry (LC-MS), followed by their identification by MS/MS. The main advantage of this pipeline over classical bottom-up proteomics (analysis of peptides generated by protein enzymatic digestion) is to directly inform on the presence of potential combinations of post-translational modifications, splicing events and/or mutations, thereby providing in-depth characterization of proteoforms (Smith *et al.*, 2013). In recent years, top-down proteomics has gained momentum (Chen *et al.*, 2018), becoming a more high-throughput (Tran *et al.*, 2011) and quantitative technique (Durbin *et al.*, 2016; Ntai *et al.*, 2014). This was allowed by the development of high-resolution mass spectrometers and algorithms allowing signal deconvolution of MS spectra along the chromatographic runs (Durbin *et al.*, 2010; Gersch *et al.*, 2015; Lee *et al.*, 2002, 2009), together with software suites

dedicated to proteoform identification (Guner *et al.*, 2014; Kou *et al.*, 2016; LeDuc *et al.*, 2004; Liu *et al.*, 2012; Park *et al.*, 2017; Tran *et al.*, 2011).

Visualization of deconvoluted LC-MS data remains a key process in top-down data inspection, and the direct comparison of several LC-MS runs reveals differences in protein footprints between samples/experimental conditions. An increasing number of top-down MS articles can be found in the literature. However, only a handful include LC-MS 2D maps produced with in-house tools. These include publications from the groups of Kelleher (Durbin *et al.*, 2010; Tran *et al.*, 2011), Patrie (Plymire *et al.*, 2017), and Smith (Lee *et al.*, 2002). We published several papers presenting such 2D maps using RoWinPro deconvolution (Gersch *et al.*, 2015) plotted with Excel (Carel *et al.*, 2017; Parra *et al.*, 2017), but this workflow is very low throughput and does not facilitate dynamic data inspection. For these reasons, downloadable



**Fig. 1.** Screenshot of the VisioProt-MS web application in “MS” mode. One file is uploaded (the color scale corresponds to the log<sub>10</sub>-transformed intensity), and an example of a 2-files upload is presented in the central insert (the original file and the overlaid file are in red and blue, respectively)

software suites were developed, such as ICR-2TLS, which generates LC-MS 2D maps (Lourette *et al.*, 2010; Sharma *et al.*, 2007; Tian *et al.*, 2010, 2012; Wu *et al.*, 2009). Since last year, the software suite Informed-Proteomics produces similar 2D maps (Park *et al.*, 2017; Zhou *et al.*, 2017). These provide many functionalities applied to top-down proteomics, but their visualization functions do not allow the direct comparison of several LC-MS runs on the same map.

Thus, we developed a standalone tool to visualize, inspect and compare the molecular weights (MWs) of eluting proteoforms against their retention times (RT). VisioProt-MS is a user-friendly and highly compatible open source web application that plots and overlays interactive 2D maps from deconvoluted LC-MS run(s). It is designed for dynamic data inspection as well as for creating publication quality figures. VisioProt-MS is compatible with the following bioinformatics tools: RoWinPro (Gersch *et al.*, 2015), Intact Protein Analysis (BioPharma Finder™ 3.0, Thermo Fisher Scientific), DataAnalysis™ 4.2 (Bruker) (Schmit *et al.*, 2018), TopFD (TopPIC Suite), (Kou *et al.*, 2016) and ProMex (Informed-Proteomics) (Park *et al.*, 2017) for deconvolution of LC-MS data; and ProSight PD (Proteome Discoverer, Thermo), TopPIC (TopPIC Suite, (Kou *et al.*, 2016) and MSPathFinder (Informed-Proteomics, (Park *et al.*, 2017)) for LC-MS/MS data. VisioProt-MS quickly provides an overview of all the detected MWs, reflecting data quality and reproducibility in terms of observed MWs, intensities and RT. It allows comparison of not only multiple LC-MS runs (including from different deconvolution suites), but also LC-MS and LC-MS/MS runs of the same sample. Its dynamic features enable to pinpoint potential new proteoforms, quickly reject wrongly assigned Protein Spectral Matches (PSM) and spot intense proteoforms that remain unassigned.

## 2 Materials and methods

VisioProt-MS is a Shiny-based web application implemented in R (3.4.2), using the RStudio environment. It depends on the shiny (1.0.3); ggplot2 (2.2.1); Plotly (4.7.1); dplyr (0.7.2); data table (1.10.4-3); shinyBS (0.61) and RcolorBrewer (1.1-2) packages. R and Shiny (1.5.5.872) are hosted by a virtual Linux platform running on Ubuntu (16.04.3).

## 3 Results

VisioProt-MS enables users with no programming skills to generate, visualize and analyze 2D maps of intact protein LC-MS runs, and to overlay the corresponding identifications from top-down analysis. It is organized in two inter-dependent panels that are presented in Figure 1: the right panel displays the output plot generated in real-time from the parameters defined in the left panel. The user can navigate between two visualization modes: “MS” to inspect one or several LC-MS file(s) for comparison of biological conditions/deconvolution parameters, or for batch-to-batch comparison; and “MS/MS” to visualize results from top-down analysis. In the following paragraphs, we will go through the specific features of these two modes, and then describe their common functionalities.

### 3.1 MS

We provide examples of compatible LC-MS input formats in the help document available by clicking on the help button (“?”). When available, the entire profile of eluted species are plotted point by point to get a “chromatogram-like” figure (like in Fig. 1). Since ProMex and BioPharma Finder™ only output the start/end elution times, VisioProt-MS outputs the range of each eluting proteoform (see help document for output examples). The application will also be compatible with outputs from any upcoming sliding window deconvolution tool, provided that the input table corresponds to a format described in the help document.

Selecting one file renders a plot with color-coded intensity whereas multiple files selection (files from the same folder) allows the comparison of several MS runs. Inspection of these protein footprints enables the direct visualization of co-eluting modified or truncated species. Additionally, overlapping different LC-MS acquisitions permits comparison between different sample conditions (e.g. treated versus non-treated samples), batch-to-batch comparisons, or quality control assessment of the LC-MS system, which are facilitated by the color-coding and zooming options.

### 3.2 MS/MS

In the “MS/MS” mode, VisioProt-MS plots a 2D map of the MS/MS spectra acquired during top-down analysis. The proportion of assigned versus unassigned MS/MS directly informs the user of the overall quality

of the data. Furthermore, the RT and MWs distributions of the assigned species can highlight eventual issues with the MS/MS acquisition parameters, search parameters or chromatographic methods. Besides quality control, this mode allows the thorough inspection of top-down proteomics data. The user can hide unassigned MS/MS spectra with the option “Hide MS/MS without ID”, and localize on the map the identified proteins by selecting them in the box “Select the ID to highlight”. It is possible to overlay MS/MS data with the LC-MS trace of the same sample (from the same LC-MS/MS run or a separate LC-MS acquisition) in order to spot false positive MS/MS hits that do not overlap with a truly deconvoluted protein (*i.e.* for several seconds and not just an artefact). The user can also identify the proteoforms clearly visible in LC-MS but not identified in the LC-MS/MS analysis, which can be due to the absence of the corresponding sequence from the database, a narrow error window, poor selection or fragmentation. All these parameters can then be optimized to improve identification results.

### 3.3 Common features

VisioProt-MS makes it quick and simple to inspect intact protein LC-MS and top-down proteomics data. Test files are provided for both MS and MS/MS modes and can be loaded by clicking on the “Using test mode” checkboxes. It is easy to zoom-in using mouse selection, and to change the plot parameters with the options provided on the left panel. A thorough inspection of the deconvoluted data can be conducted when “Show data labels” is checked. Then, dynamic labels appear on mouse hovering, returning the following values from the input tables: RT (min); molecular weight of the protein (Da); signal intensity; protein identity and file name when applicable. These can be of great use to identify peaks of interest in the 2D map. Because dynamic labeling significantly slows down the application, we kept it as an option, and we advise the user to check it only after optimization of the rendering parameters (colors, point size, threshold and zooming). When comparing several LC-MS files, the “Show data labels” mode allows the user to isolate or mask files by clicking on their names in the figure legend, thereby facilitating data inspection. To quicken plot rendering, we decided to plot only 20% of the data points with the highest intensity. This parameter can be defined manually in the dedicated field. Modifying this threshold results in changes in the axis and color ranges due to the appearance/disappearance of low intensity points. A selection of color scales offers the choice between diverging or sequential palettes (see the help document to visualize the color-scales). When several files are plotted together, the colors on the plot are file-dependent, whereas when only one file is plotted, the colors correspond to the log10-transformed intensity values from the input table.

Export of the displayed map can be done by a simple click on the button “Download.pdf”, “Download.png” or “Download.svg”. By default, this output file is named “VisioProt-MS\_” followed by the name of the input file and the date. It is saved in the “Downloads” folder of the user computer, and can be used as it is for publication purposes.

All the functionalities of VisioProt-MS are described in details in the help webpage available by clicking on the help button (“?”).

### Acknowledgements

The authors would like to thank S. N'Diaye (Thermo Fisher Scientific), J. Vialaret (IRMB, France) and P.O. Schmit (Bruker Daltonique) for sending us example files.

### Funding

This project was supported by the Région Occitanie and the French Ministry of Research with the Investissement d'Avenir Infrastructures Nationales

en Biologie et Santé program (ProFI, Proteomics French Infrastructure project, ANR-10-INBS-08).

*Conflict of Interest:* none declared.

### References

- Carel, C. *et al.* (2017) Identification of specific posttranslational O-mycoloylations mediating protein targeting to the mycomembrane. *Proc. Natl. Acad. Sci.*, **114**, 4231–4236.
- Chen, B. *et al.* (2018) Top-down proteomics: ready for Prime Time? *Anal. Chem.*, **90**, 110–127.
- Durbin, K.R. *et al.* (2010) Intact mass detection, interpretation, and visualization to automate Top-Down proteomics on a large scale. *Proteomics*, **10**, 3589–3597.
- Durbin, K.R. *et al.* (2016) Quantitation and identification of thousands of human proteoforms below 30 kDa. *J. Prot. Res.*, **15**, 976–982.
- Gersch, M. *et al.* (2015) A mass spectrometry platform for a streamlined investigation of proteasome integrity, posttranslational modifications, and inhibitor binding. *Chem. Biol.*, **22**, 404–411.
- Guner, H. *et al.* (2014) MASH Suite: a user-friendly and versatile software interface for high-resolution mass spectrometry data interpretation and visualization. *J. Am. Soc. Mass. Spectrom.*, **25**, 464–470.
- Kou, Q. *et al.* (2016) TopPIC: a software tool for top-down mass spectrometry-based proteoform identification and characterization. *Bioinformatics*, **32**, 3495–3497.
- LeDuc, R.D. *et al.* (2004) ProSight PTM: an integrated environment for protein identification and characterization by top-down mass spectrometry. *Nucleic Acids Res.*, **32**, W340–W345.
- Lee, J.E. *et al.* (2009) A robust two-dimensional separation for top-down tandem mass spectrometry of the low-mass proteome. *J. Am. Soc. Mass Spectrom.*, **20**, 2183–2191.
- Lee, S.-W. *et al.* (2002) Direct mass spectrometric analysis of intact proteins of the yeast large ribosomal subunit using capillary LC/FTICR. *Proc. Natl. Acad. Sci.*, **99**, 5942–5947.
- Liu, X. *et al.* (2012) Protein identification using top-down spectra. *Mol Cell Proteomics*, **11**, M111.008524.
- Lourette, N. *et al.* (2010) A top-down LC-FTICR MS-based strategy for characterizing oxidized calmodulin in activated macrophages. *J. Am. Soc. Mass Spectrom.*, **21**, 930–939.
- Ntai, I. *et al.* (2014) Applying label-free quantitation to top down proteomics. *Anal. Chem.*, **86**, 4961–4968.
- Park, J. *et al.* (2017) Informed-proteomics: open-source software package for top-down proteomics. *Nat. Meth.*, **14**, 909–914.
- Parra, J. *et al.* (2017) Scrutiny of *Mycobacterium tuberculosis* 19 kDa antigen proteoforms provides new insights in the lipoglycoprotein biogenesis paradigm. *Sci. Rep.*, **7**, 43682.
- Plymire, D.A. *et al.* (2017) Continuous elution proteoform identification of myelin basic protein by superficially porous reversed-phase liquid chromatography and fourier transform mass spectrometry. *Anal. Chem.*, **89**, 12030–12038.
- Schmit, P.-O. *et al.* (2018) Towards a routine application of Top-Down approaches for label-free discovery workflows. *J. Prot.*, **175**, 12–26.
- Sharma, S. *et al.* (2007) Proteomic profiling of intact proteins using WAX-RPLC 2-D separations and FTICR mass spectrometry. *J. Prot. Res.*, **6**, 602–610.
- Smith, L.M. *et al.* (2013) Proteoform: a single term describing protein complexity. *Nat Meth.*, **10**, 186–187.
- Tian, Z. *et al.* (2012) Enhanced top-down characterization of histone post-translational modifications. *Genome Biol.*, **13**, R86.
- Tian, Z. *et al.* (2010) Two-dimensional liquid chromatography system for on-line top-down mass spectrometry. *Proteomics*, **10**, 3610–3620.
- Tran, J.C. *et al.* (2011) Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature*, **480**, 254–258.
- Wu, S. *et al.* (2009) An integrated top-down and bottom-up strategy for broadly characterizing protein isoforms and modifications. *J. Prot. Res.*, **8**, 1347–1357.
- Zhou, M. *et al.* (2017) Profiling of histone post-translational modifications in mouse brain with high-resolution top-down mass spectrometry. *J. Prot. Res.*, **16**, 599–608.