



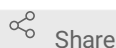
Aug 12, 2022

Image Visualization and Proteoform Assignment of MALDI-MSI from LCMS Experimental Databases

David J Degnan¹, Kevin J. Zemaitis¹, Dusan Velickovic¹, Mowei Zhou¹, Ljiljana.PasaTolic¹

¹Pacific Northwest National Laboratory

1 Works for me



Share

This protocol is published without a DOI.

Human BioMolecular Atlas Program (HuBMAP) Method Development Community

PNNL TTD



Kevin J. Zemaitis

ABSTRACT

Scope:

The protocol describes a workflow for processing mass spectrometry imaging (MSI) data of high resolution intact proteins/proteoforms by MALDI. It involves the visualization and image generation using commercial software and also peak annotation using open source code.

Expected Outcomes:

Tabular output of matched proteoforms, as well as a trelliscope display of overlaid isotopic distributions allowing for the annotation of high-resolution accurate mass distributions of proteoforms.

PROTOCOL CITATION

David J Degnan, Kevin J. Zemaitis, Dusan Velickovic, Mowei Zhou, Ljiljana.PasaTolic 2022. Image Visualization and Proteoform Assignment of MALDI-MSI from LCMS Experimental Databases. **protocols.io**
<https://protocols.io/view/image-visualization-and-proteoform-assignment-of-m-cewjtfcn>



FUNDERS ACKNOWLEDGEMENT

National Institutes of Health (NIH) Common Fund, Human Biomolecular Atlas Program (HuBMAP)
Grant ID: UG3CA256959-01

LICENSE

————— This is an open access protocol distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

CREATED

Aug 05, 2022

LAST MODIFIED

Aug 12, 2022

OWNERSHIP HISTORY

Aug 05, 2022  Mowei Zhou Pacific Northwest National Laboratory

Aug 12, 2022  Kevin J. Zemaitis

PROTOCOL INTEGER ID

68267

BEFORE STARTING

Both MALDI imaging and TDP by LCMS/MS needs to be completed and processed per other outlined protocols, packages need to be ready to use on desktop or laptop as well.

Data pre-processing

- 1 The .xml file from the MALDI source registering pixel coordinates is renamed to that of the .RAW file generated from the instrument, both files are then transferred to a separate imaging workstation after being backed up to an internal database.
- 2 These files are then imported into SCiLS Lab Pro (v.2021c) for preliminary visualization.
 - 2.1 Binning parameters within other versions have been noted to alter isotopic distributions and over- or under-bin peaks over the broad mass range, automatic parameters within 2021c were found to be ideal in most cases, as the peak width within these analyses change over the broad range of mass-to-charge values.
 - 2.2 For consistency, peak lists are imported into SCiLS Lab to export a .imzML for ingestion.

Proteoform image visualization

- 3 Proteoforms are visualized from the most abundant isotopologue, the monoisotopic mass from the centroid of this peak is used for determining the mass error of the annotated proteoform.

Proteoform assignment via ProteoMatch

- 4 Next, the ProteoMatch tool is utilized to calculate and match isotoping profiles to spectra. Below is a general description of the pipeline:
 1. *Required*: Calculate molecular formulas and mass shifts with `calculate_molform()`
 2. *Optional*: Filter noisy peaks and the peak MZ range with `filter_peaks()`
 3. *Required*: Match reference isotope profiles to experimental data with `match_proteoform_to_ms1()`. Protein sequences without PTMs are accepted as well.
 4. *Optional*: Visualize results with `plot_Ms1Match()` or `proteomatch_trelliscope()`.

ProteoMatch

[source](#) by David Degnan

Note: This utilizes spatially resolved LCM-TDP obtained via LCMS/MS, input files for experimental databases are outlined in other protocols and need to be processed prior to completing this annotation.

A main pipeline function called `run_proteomatch()` can be used, and requires three files, which are all described below:

- A proteoform .csv file (see section 4.1)
- A .mzML file (see section 4.2)
- A .xlsx settings file (see section 4.3)

- 4.1 Prepare a .csv proteoform file with a “Proteoform” and “Protein” (any string). Proteoform annotation generally follow the ProForma convention.

LeDuc RD, Schwämmle V, Shortreed MR, Cesnik AJ, Solntsev SK, Shaw JB, Martin MJ, Vizcaino JA, Alpi E, Danis P, Kelleher NL, Smith LM, Ge Y, Agar JN, Chamot-Rooke J, Loo JA, Pasa-Tolic L, Tsybin YO (2018). ProForma: A Standard Proteoform Notation.. Journal of Proteome Research.
<https://doi.org/10.1021/acs.jproteome.7b00851>

Post-translation modifications (PTMs) can be annotated either by name

(UniMod definition) or mass shifts. An example proteoform annotation:

```
"M. (S) [Acetyl]GRGKGGKGLGKGGAKRHRK (VLRDNIQGITKPAIRRLAR)
[28.0315]RGGVKRISGLIYEETRGVLKVFLENVIRDAVTYTEHAKRKTVTAMDV
VYALKRQGRTLYGFGG. "
```

The periods are marking the starting and ending residues of the proteoform, which is used by TopPIC. The residues left to the first period and right to the second period (if any) are removed in the formula generation. Custom definition can be created by updating the backend glossary on github (ProteoMatch/inst/extdata/Unimod.csv) or by submitting an issue request on the github page.

- 4.2 Prepare a .mzML file by converting the .RAW with MSConvert and the `peak picking -1` flag enabled.
- 4.3 A .xlsx settings file will also need to be provided. For a template, see github (ProteoMatch/inst/extdata/Defaults.xlsx).
- 4.4 The last parameter in the run_proteomatch() function is to indicate the output directory. Note that the main pipeline function run_proteomatch() will execute only if the correct files are added.
- 4.5 As the pipeline runs, the following files will be generated: a .csv with molecular formulas, a .csv with filtered peaks, a .csv with matched peaks, and the trelliscope display of best matches (Pearson Correlation ≥ 0.7). This value can be modified in the settings .xlsx file row "CorrelationMinimum." Further details about each of the functions can be explored in the R documentation or by exploring the ProteoMatch vignettes (using the Vignettes() function).