

IsoMatchMS: Open-Source Software for Automated Annotation and Visualization of High Resolution MALDI-MS Spectra

David J. Degnan^{1,*}, Kevin J. Zemaitis², Logan A. Lewis¹, Lee Ann McCue¹, Lisa M. Bramer¹, James M. Fulcher², Dušan Veličković², Ljiljana Paša-Tolić², Mowei Zhou^{2,*}

¹Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA, 99354, USA, and ²Environmental Molecular Sciences Division, Pacific Northwest National Laboratory, Richland, WA, 99354, USA.

*David J. Degnan: david.degnan@pnnl.gov and Mowei Zhou: mowei.zhou@pnnl.gov

Mass spectrometry. Isotope profile. MALDI-MS. Trelliscope. R package.

ABSTRACT: Due to its speed, accuracy, and adaptability to various sample types, matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) has become a popular method to identify molecular isotope profiles from biological samples. Often MALDI-MS data does not include tandem MS fragmentation data, and thus the identification of compounds in samples requires external databases so that the accurate mass of detected signals can be matched to known molecular compounds. Most relevant software tools are focused on small molecules (*e.g.*, metabolites, lipids), and cannot be easily adapted to protein data due to their more complex isotopic distributions. Here, we present an R package called *IsoMatchMS* for the automated annotation of MALDI-MS data for multiple datatypes (*e.g.*, intact proteins, lipids, metabolites, etc.). This tool accepts already computed molecular formulas, or for proteomics applications, can compute molecular formulas from a list of input peptides or proteins including proteins with post-translational modifications. Visualization of all matched isotopic profiles are provided in a highly accessible HTML format called a trelliscope display, which allows users to filter and sort by several parameters such as match scores and the number of peaks matched. *IsoMatchMS* simplifies the annotation and visualization of MALDI-MS data for downstream analyses.

INTRODUCTION

MALDI-MS (matrix-assisted laser desorption/ionization mass spectrometry) is a popular technology for characterizing molecular compounds and their location within biological systems (*e.g.*, tissues, microbial colonies, etc.) due to its high sensitivity and high throughput, relative to other spatially resolved MS technologies.^{1,2} In a typical workflow, peak assignments are based on matching the experimental masses to databases of masses for known molecules, either from parallel experiments or *in silico* prediction. To precisely match an entry in one of these databases, high resolution data with isotopic profiles are preferable for high confidence identifications. Of particular interest for MALDI-MS analysis are proteins and their modified forms (proteoforms), which have roles in epigenetics (gene regulation), cell signaling, protein degradation, and signal transduction.^{3,4} However, only recently have developments addressing ion transmission and resolving power limitations⁵⁻⁹ enabled isotopically resolved MALDI-MS spectra for intact proteins (~20 kDa). Therefore, most existing software for annotating MALDI-MS data is largely focused on molecules with relatively small isotopic profiles (*e.g.* lipids metabolites),¹⁰⁻¹² with limited support for peptides (digested proteins).¹³⁻¹⁶ Consequently, users interested in peptide and protein MALDI-MS data often must manually combine the capabilities of several different tools for peak assignment, which is a time-consuming process.

Here, we developed an R package called *IsoMatchMS* (<https://github.com/PNNL-HubMAP-Proteoform-Suite/IsoMatchMS>) to support both the analysis of MALDI-MS data with molecular formula annotation, and the analysis of intact protein and peptide MALDI-MS data from ProForma¹⁷ strings. *IsoMatchMS* derives molecular formulas from ProForma¹⁷ strings or the modification formats from five different proteomics identification tools: MSPath-Finder¹⁴, ProSightPC¹⁸, pTop¹⁹, TopPIC²⁰, and MS-GF+.²¹ Then, *IsoMatchMS* calculates full theoretical isotope profiles, matches them to a summed spectrum, and visualizes the overlap between the two in an HTML display (called a “trelliscope²² display”) where each plot is an overlay of a single isotope profile for a molecule on the summed experimental spectra. Users can then sort the display by high-scoring distributions to confirm annotations. Unique to existing open-source software tools with functions to identify high-quality annotations,¹⁰⁻¹⁶ *IsoMatchMS* supports spectral summing, common proteomics datatypes (both peptide and intact protein), complex compounds with modifications and unknown mass shifts, any adducts with a known mass, and visualizes the results in highly shareable and sortable trelliscope displays. By unifying the analysis process for multiple types of biological molecules into a singular pipeline, *IsoMatchMS* reduces the time and effort for the annotation of MALDI-MS data.

EXPERIMENTAL SECTION

Package Design. *IsoMatchMS* was written in R²³ 4.2.2 and has a main pipeline function which requires three inputs: 1) molecular formulas or ProForma¹⁷ strings, 2) MS peak data, and 3) a settings file (an excel .xlsx). The molecular formulas or ProForma¹⁷ strings need only be a vector of characters in R and thus can be read from any file format that can be read into R. Similarly, the MS peak data is a *pspecterlib*²⁴ `peak_data` object that can be created from any two numeric vectors (*m/z* and intensity) or extracted from a `mzML` or a ThermoFisher raw file with a *pspecterlib*²⁴ wrapper function, which uses *mzR*²⁵ and *rawrr*.²⁶ The settings file specifies several parameters such as *m/z* range and *m/z* error; examples are provided within the package. The run `isomatchms()` function uses these objects to automatically perform the isotope calculations, peak matching, and trelliscope display generation.

The ProForma¹⁷ strings can originate from experimental (e.g., LC-MS/MS) or protein sequence (e.g., UniProt²⁷) databases. If ProForma strings are not provided by an identification tool such as TopPIC²⁰, *IsoMatchMS* contains a function to derive them from `mzid` files generated by tools such as MS-GF⁺²¹ or from various modification formats, including those used in the tools MSPathFinder¹⁴, ProSightPC¹⁸, and pTop.¹⁹ Molecular formulas are also accepted in lieu of ProForma strings, written simply as element and number of atoms with no spaces (e.g., “C2H5NO2”). MS1 spectra can be summed with other software (e.g., instrument vendor software), or calculated from `mzML` files with *IsoMatchMS*. Within the settings files, users can set a variety of parameters to optimize performance, including the isotope algorithm to use (either *Rdisop*²⁸ or *isopat*²⁹), range of *m/z* values to search, an abundance noise filter, a minimum number of isotopic peaks to identify, a maximum number of charge states to investigate, and PPM and abundance tolerances to be considered a “match.” The example settings files in *IsoMatchMS* contain more details about each parameter, along with suggested defaults. For intact protein datasets, we suggest a minimum of 5 identified isotopes, higher noise filters (2.5-10%), and using the *Rdisop*²⁸ isotoping algorithm due to its higher precision at heavier masses. For smaller isotope profiles, we suggest a minimum of 2 identified isotopes and lower noise filters (0-2.5%). Available isotope scores include

Pearson correlation and the absolute relative error (1) where A_M is the measured abundance, A_C is the calculated abundance, and n is the number of peaks.

$$\frac{1}{n} * \sum \frac{|A_M - A_C|}{A_M}$$

(1)

For higher mass profiles (intact proteomics), we suggest a Pearson correlation of 0.9 or greater, and 0.7 or greater otherwise. Alongside the trelliscope display, the main pipeline exports `csv` files with molecular formulas and masses, all matched isotope peaks regardless of score, and the processed mass spectra.

MS Data. MALDI-MS data from intact protein and tryptic peptide analyses were collected from sections of a fresh-frozen and non-embedded female Sprague Dawley rat brains (*Rattus norvegicus*).^{30,31} Briefly, tissues were purchased from BioIVT (Westbury, NY) and cryosectioned as 10 μm sections onto either ITO (indium-tin oxide) for MALDI-MS or PEN (polyethylene naphthalate)-membrane slides for laser capture microdissection. Briefly, intact proteins were analyzed after a tissue fixation and lipid depletion with a protocol, and tryptic peptides were analyzed after *in-situ* deposition of trypsin directly on the tissue and overnight incubation, this is described in great depth elsewhere.^{30,32} Glycan data was collected from formalin-fixed paraffin-embedded (FFPE) soybean nodules (*Glycine max* cv. Williams 82). Briefly, FFPE soybean nodules were cryosectioned as 7 μm sections onto ITO slides, where PNGase was applied and incubated using a protocol described in greater depth elsewhere.³³ Both glycans and intact proteins were analyzed on a custom UHRM Q Exactive HF Orbitrap (Thermo Scientific, Bremen, DE) fitted with an elevated pressure MALDI source (Spectroglyph, Kennewick, WA) operated under a custom privilege license.³² Tryptic peptides were analyzed on a 12T solariX Fourier transform ion cyclotron resonance mass spectrometer (Bruker Daltonics, Bremen, DE). All raw MALDI-MS data were summed with Mozaic (v.2022.5.1.b1; Spectroswiss, Lausanne, CH). Experimental proteoform/peptide databases were based on LC-MS/MS data from published results of intact³⁰ and digested³¹ LC-MS/MS. All MS Data is included within the R package.

Table 1. *IsoMatchMS* performance on intact protein, peptide, and glycan datasets.

Dataset	Intact Protein	Peptide	Glycans
Molecular formulas searched	2125	2594	3667
Adducts	Proton	Proton	Proton, Sodium
Number of manually verified high confidence annotations	21	28	26
Number of manual identifications also identified by <i>IsoMatchMS</i> at a Pearson correlation score of 0.70	21 / 21	28 / 28	26 / 26

Number of manual identifications also identified by <i>IsoMatchMS</i> at a Pearson correlation score of 0.90	19 / 21	27 / 28	21 / 26
Number of manual identifications also identified by <i>IsoMatchMS</i> at a Pearson correlation score of 0.95	15 / 21	27 / 28	18 / 26
Seconds to complete using an Apple M1 Chip with 10 cores	72	54	33

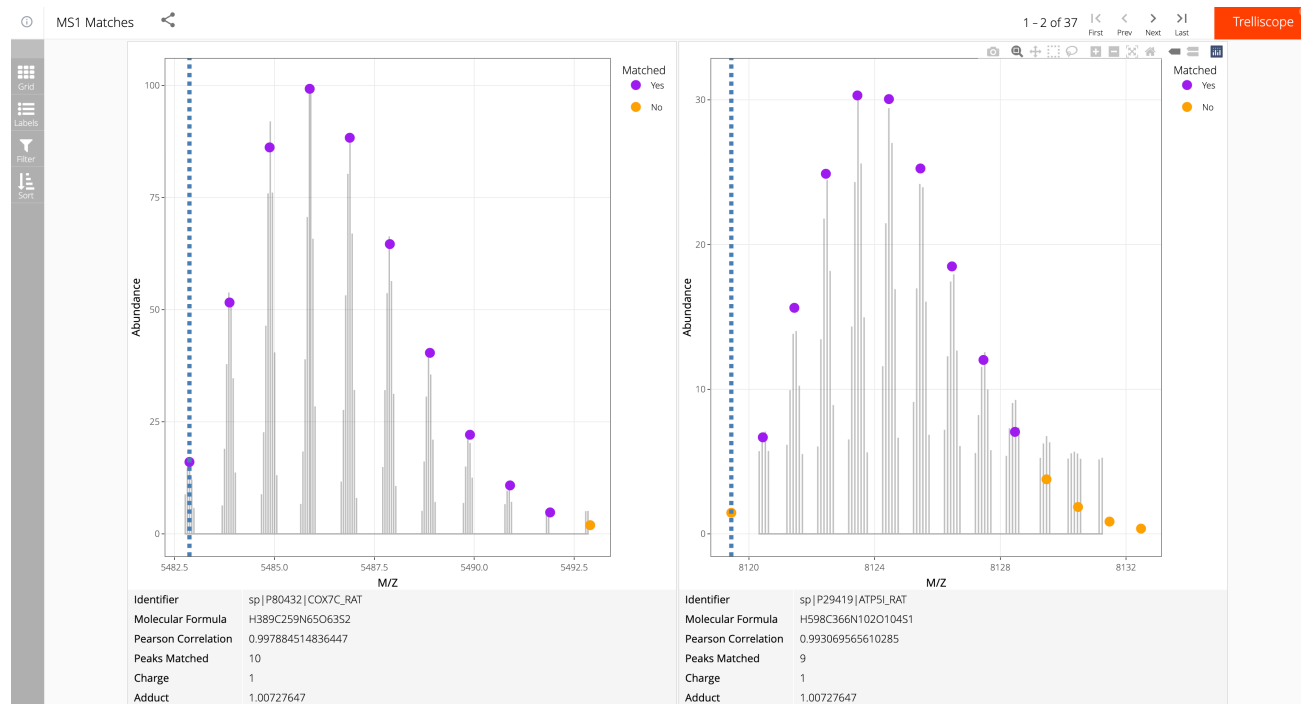


Figure 1. Trelliscope display of the isotopic distributions of two relevant proteoforms from the intact protein dataset. The blue line is the monoisotopic mass, and the black lines are the experimental mass spectra. The points represent the theoretical isotope distribution where the purple points are fit isotopes, and the orange are not.

RESULTS AND DISCUSSION

For each of the three tested datasets (intact protein, peptide, and glycan data), *IsoMatchMS* calculated, matched, and visualized isotope profiles in under 90 seconds (Table 1). Datasets where molecular formula information was already calculated from a theoretical database (glycans) completed faster than formulas derived from experimental ProForma¹⁷ strings. Across all datasets, highly confident annotations derived by using other tools^{12, 14} were also identified by *IsoMatchMS* at a Pearson score threshold of 0.7 (Table 1). Even at higher Pearson score thresholds of 0.9 and 0.95, there was high overlap between the number of manual and *IsoMatchMS* annotations. An example of the trelliscope display output for the intact protein data was provided (Fig. 1). While *IsoMatchMS* was able to annotate several dozen high-confidence isotopic profiles within the presented datasets, the number of matches per number of molecular formulae searched (Table 1) was relatively low. This was due to the reduced dynamic range of MALDI-MS detect several dozen to a couple hundred species in comparison to LC-MS/MS which can routinely detect several hundred intact proteins or several thousand digested peptides.

While most intact proteins detected by MALDI-MS can be readily assigned from experiment intact protein LC-MS/MS databases, the proteome coverage will never perfectly align even in the case of the best experimental databases generated from serial sections of tissue.

Regardless, manual assignment and validation of molecular isotopic envelopes can take up to several hours or more per analysis and require the use of multiple tools depending on the specific restrictions per tool and the aims of the analysis. Centralizing the steps of the bioinformatics workflow, which includes sequence to molecular formula conversion, the addition of known and unknown mass shifts, and isotopic matching, within *IsoMatchMS* allows for a preliminary analysis of high-resolution MALDI-MS with an experimental or protein sequence database within minutes. This improvement in speed allows for more high-throughput analyses of MALDI-MS data and reduces the knowledge burden associated with using several tools to process data for different types of molecules. While the curation of an experimental or protein sequence database still requires sample dependent supplemental efforts, this tool vastly improves the throughput and reproducibility of MALDI-MS workflows for biomolecules. As more

experimental data become available in public repositories, scoring matches may be improved to the point of fully automated peak annotations.

CONCLUSION

IsoMatchMS is an open-source R package for identifying molecular annotations from high resolution MALDI-MS data. Trelliscope displays allow users to quickly sort through results to select interesting annotations for further study in downstream analyses, including differential abundance statistics or mass spectrometry imaging. Overall, *IsoMatchMS* reduces the time burden for identifying high quality annotations, and is an example of a flexible tool built for multiple biological molecules (e.g., metabolites, intact or digested proteins, lipids). *IsoMatchMS* provides a critical first screening effort for MALDI-MS data, but manual verification of annotations in the trelliscope display should still occur as settings are variable.

AUTHOR INFORMATION

Author Contributions

LPT and MZ conceptualized and managed the project; KJZ, LMB, JMF, DV, and MZ contributed to the tool and its design; LAM conceptualized the software and supervised development; DJD designed the tool; LAL contributed to the code and documentation; and DJD and KJZ wrote the first draft of the manuscript. All authors have edited and given approval to the final version of the manuscript.

Funding Sources

This research was funded by the National Institutes of Health (NIH) Common Fund, Human Biomolecular Atlas Program (HuBMAP) grant UG3CA256959-01. A portion of this work was performed under project doi.org/10.46936/staf.proj.2020.51770/60000309 (LPT) at the Environmental Molecular Science Laboratory (EMSL), a DOE Office of Science User Facility sponsored by the Biological and Environmental Research program under Contract No. DE-AC05-76RL01830. The authors declare no competing financial interest.

ACKNOWLEDGEMENTS

We thank Camryn Pettenger-Wiley for her assistance in acquiring the peptide dataset.

REFERENCES

- (1) Taylor, M. J.; Lukowski, J. K.; Anderton, C. R. Spatially Resolved Mass Spectrometry at the Single Cell: Recent Innovations in Proteomics and Metabolomics. *J Am Soc Mass Spectrom* **2021**, *32* (4), 872-894. DOI: 10.1021/jasms.0c00439.
- (2) Singhal, N.; Kumar, M.; Kanaujia, P. K.; Virdi, J. S. MALDI-TOF mass spectrometry: an emerging technology for microbial identification and diagnosis. *Front Microbiol* **2015**, *6*, 791. DOI: 10.3389/fmicb.2015.00791.
- (3) Forgrave, L. M.; Wang, M.; Yang, D.; DeMarco, M. L. Proteoforms and their expanding role in laboratory medicine. *Pract. Lab. Med.* **2022**, *28*, e00260, DOI: 10.1016/j.plabm.2021.e00260.
- (4) Lee, M. J.; Yaffe, M. B. Protein Regulation in Signal Transduction. *Cold Spring Harb. Perspect. Biol.* **2016**, *8*(6), DOI: 10.1101/cshperspect.a005918.
- (5) Zemaitis, K. J.; Velickovic, D.; Kew, W.; Fort, K. L.; Reinhardt-Szyba, M.; Pamreddy, A.; Ding, Y.; Kaushik, D.; Sharma, K.; Makarov, A. A.; et al. Enhanced Spatial Mapping of Histone Proteoforms in Human Kidney Through MALDI-MSI by High-Field

UHMR-Orbitrap Detection. *Anal Chem* **2022**, *94* (37), 12604-12613. DOI: 10.1021/acs.analchem.2c01034.

- (6) Spraggins, J. M.; Rizzo, D. G.; Moore, J. L.; Rose, K. L.; Hammer, N. D.; Skaar, E. P.; Caprioli, R. M. MALDI FTICR IMS of Intact Proteins: Using Mass Accuracy to Link Protein Images with Proteomics Data. *J Am Soc Mass Spectrom* **2015**, *26* (6), 974-985. DOI: 10.1007/s13361-015-1147-5.

- (7) Nicolardi, S.; Switzar, L.; Deelder, A. M.; Palmblad, M.; van der Burgt, Y. E. Top-down MALDI-in-source decay-FTICR mass spectrometry of isotopically resolved proteins. *Anal Chem* **2015**, *87* (6), 3429-3437. DOI: 10.1021/ac504708y.

- (8) Prentice, B. M.; Ryan, D. J.; Van de Plas, R.; Caprioli, R. M.; Spraggins, J. M. Enhanced Ion Transmission Efficiency up to m/z 24 000 for MALDI Protein Imaging Mass Spectrometry. *Anal Chem* **2018**, *90*(8), 5090-5099. DOI: 10.1021/acs.analchem.7b05105.

- (9) Han, J.; Permentier, H.; Bischoff, R.; Groothuis, G.; Casini, A.; Horvatovich, P. Imaging of protein distribution in tissues using mass spectrometry: An interdisciplinary challenge. *TrAC Trends in Anal Chem* **2019**, *112*, 13-28.

- (10) Tortorella, S.; Tiberi, P.; Bowman, A. P.; Claes, B. S. R.; Scupakova, K.; Heeren, R. M. A.; Ellis, S. R.; Cruciani, G. LipostarMSI: Comprehensive, Vendor-Neutral Software for Visualization, Data Analysis, and Automated Molecular Identification in Mass Spectrometry Imaging. *J Am Soc Mass Spectrom* **2020**, *31*(1), 155-163. DOI: 10.1021/jasms.9b00034.

- (11) Goracci, L.; Tortorella, S.; Tiberi, P.; Pellegrino, R. M.; Di Veroli, A.; Valeri, A.; Cruciani, G. Lipostar, a Comprehensive Platform-Neutral Cheminformatics Tool for Lipidomics. *Anal Chem* **2017**, *89*(11), 6257-6264. DOI: 10.1021/acs.analchem.7b01259.

- (12) Palmer, A.; Phapale P.; Chernyavsky, I.; Lavigne, R.; Fay, D.; Tarasov, A.; Kovalev, V.; Fuchser, J.; Nikolenko, S.; Pineau, C.; Becker, M.; Alexandrov, T. FDR-controlled metabolite annotation for high-resolution imaging mass spectroemtry. *Nat Methods* **2017**, *14*, 57-60.

- (13) Greer, J. B.; Early, B. P.; Durbin, K. R.; Patrie, S. M.; Thomas, P. M.; Kelleher, N. L.; LeDuc, R. D.; Fellers, R. T. ProSight Annotator: Complete control and customization of protein entries in UniProt XML files. *Proteomics* **2022**, *22*(11-12), e2100209. DOI: 10.1002/pmic.202100209.

- (14) Park, J.; Piehowski, P. D.; Wilkins, C.; Zhou, M.; Mendoza, J.; Fujimoto, G. M.; Gibbons, B. C.; Shaw, J. B.; Shen, Y.; Shukla, A. K.; et al. Informed-Proteomics: open-source software package for top-down proteomics. *Nat Methods* **2017**, *14*(9), 909-914. DOI: 10.1038/nmeth.4388.

- (15) Fornelli, L.; Srzentic, K.; Huguet, R.; Mullen, C.; Sharma, S.; Zabrouskov, V.; Fellers, R. T.; Durbin, K. R.; Compton, P. D.; Kelleher, N. L. Accurate Sequence Analysis of a Monoclonal Antibody by Top-Down and Middle-Down Orbitrap Mass Spectrometry Applying Multiple Ion Activation Techniques. *Anal Chem* **2018**, *90*(14), 8421-8429. DOI: 10.1021/acs.analchem.8b00984.

- (16) Guo, G.; Papanicolaou, M.; Demarais, N. J.; Wang, Z.; Schey, K. L.; Timpson, P.; Cox, T. R.; Grey, A. C. Automated annotation and visualisation of high-resolution spatial proteomic mass spectrometry imaging data using HIT-MAP. *Nat Commun* **2021**, *12*(1), 3241. DOI: 10.1038/s41467-021-23461-w.

- (17) LeDuc, R. D.; Schwämmle, V.; Shortreed, M. R.; Cesnik, A. J.; Solntsev, S. K.; Shaw, J. B.; Martin, M. J.; Vizcaino, J. A.; Alpi, E.; Danis, P. ProForma: a standard proteoform notation. *J. Proteome Res.* **2018**, *17*(3), 1321-1325.

- (18) Zamdborg, L.; LeDuc, R. D.; Glowacz, K. J.; Kim, Y. B.; Viswanathan, V.; Spaulding, I. T.; Early, B. P.; Bluhm, E. J.; Babai, S.; Kelleher, N. L. ProSight PTM 2.0: improved protein identification and characterization for top down mass spectrometry. *Nucleic Acids Res* **2007**, *35*, W701-706. DOI: 10.1093/nar/gkm37.

- (19) Sun, R. X.; Wang, R. M.; Luo, L.; Liu, C.; Chi, H.; Zeng, W. F.; He, S. M. Accurate Proteoform Identification and Quantitation Using pTop 2.0. *Methods Mol Biol* **2022**, *2500*, 105-129. DOI: 10.1007/978-1-0716-2325-1_9.

- (20) Kou, Q.; Xun, L.; Liu, X. TopPIC: a software tool for top-down mass spectrometry-based proteoform identification and characterization. *Bioinformatics* **2016**, *32*(22), 3495-3497. DOI: 10.1093/bioinformatics/btw398.
- (21) Kim, S.; Pevzner, P. A. MS-GF+ makes progress towards a universal database search tool for proteomics. *Nat Commun* **2014**, *5*, 5277. DOI: 10.1038/ncomms6277.
- (22) Hafen, R.; Shloerke, B. trelliscope.js: Creative Interactive Trelliscope Displays. *R package version 0.2.6*, **2021**. <https://CRAN.R-project.org/package=trelliscopejs> (accessed 2023-05-01)
- (23) Ihaka, R.; Gentleman, R. R: a language for data analysis and graphics. *JCGS* **1996**, *5*(3), 299-314.
- (24) Degnan, D. J.; Bramer, L. M.; White, A. M.; Zhou, M.; Bilbao, A.; McCue, L. A. PSpectreR: A User-Friendly and Interactive Application for Visualizing Top-Down and Bottom-Up Proteomics Data in R. *J Proteome Res* **2021**, *20*(4), 2014-2020. DOI: 10.1021/acs.jproteome.0c00857.
- (25) Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; et al. A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol* **2012**, *30*(10), 918-920. DOI: 10.1038/nbt.2377.
- (26) Kockmann, T.; Panse, C. The rawrr R Package: Direct Access to Orbitrap Data and Beyond. *J Proteome Res* **2021**, *20*(4), 2028-2034. DOI: 10.1021/acs.jproteome.0c00866.
- (27) UniProt, C. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res* **2023**, *51*(D1), D523-D531. DOI: 10.1093/nar/gkac1052.
- (28) Böcker, S.; Letzel, M. C.; Lipták, Z.; Pervukhin, A. SIRIUS: decomposing isotope patterns for metabolite identification. *Bioinformatics* **2009**, *25*(2), 218-224.
- (29) Loos, M. isopat: Calculation of isotopic pattern for a given molecular formula. *R package version 1.0*, **2012**. <https://CRAN.R-project.org/package=isopat> (accessed 2023-05-01)
- (30) Liao, Y. C.; Fulcher, J. M.; Degnan, D. J.; Williams, S. M.; Bramer, L. M.; Velickovic, D.; Zemaitis, K. J.; Velickovic, M.; Sontag, R.; Moore, R. J.; et al. Spatially resolved top-down proteomics of tissue sections based on a microfluidic nanodroplet sample preparation platform. *Mol Cell Proteomics* **2023**, 100491. DOI: 10.1016/j.mcpro.2022.100491.
- (31) Zhu, Y.; Dou, M.; Piehowski, P. D.; Liang, Y.; Wang, F.; Chu, R. K.; Chrisler, W. B.; Smith, J. N.; Schwarz, K. C.; Shen, Y.; et al. Spatially Resolved Proteome Mapping of Laser Capture Microdissected Tissue with Automated Sample Transfer to Nanodroplets. *Mol Cell Proteomics* **2018**, *17*(9), 1864-1874. DOI: 10.1074/mcp.TIR118.000686.
- (32) Zemaitis, K. J.; Zhou, M.; Kew, W.; Pasa-Tolic, L. 193 nm Ultraviolet Photodissociation for the Characterization of Singly Charged Proteoforms Generated by MALDI. *J Am Soc Mass Spectrom* **2023**, *34*(2), 328-332. DOI: 10.1021/jasms.2c00302.
- (33) Velickovic, D.; Liao, Y. C.; Thibert, S.; Velickovic, M.; Anderton, C.; Voglmeir, J.; Stacey, G.; Zhou, M. Spatial Mapping of Plant N-Glycosylation Cellular Heterogeneity Inside Soybean Root Nodules Provided Insights Into Legume-Rhizobia Symbiosis. *Front Plant Sci* **2022**, *13*, 869281. DOI: 10.3389/fpls.2022.869281.

