ACS Partner Journal

Journal of the American Society for Mass Spectrometry

pubs.acs.org/jasms

Application Note

PatternLab V Handles Multiplex Spectra in Shotgun Proteomic Searches and Increases Identification

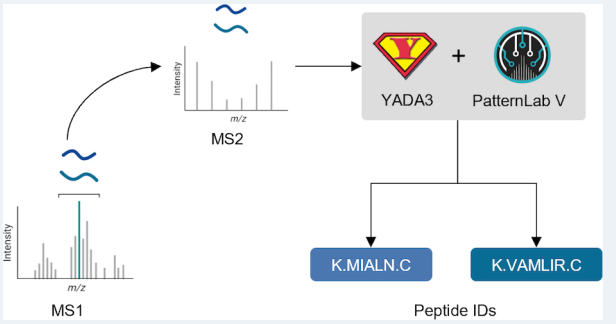
Milan A. Clasen,[+] Marlon D. M. Santos,[+] Louise Ulrich Kurt, Juliana Fischer, Amanda C. Camillo-Andrade, Lucas Albuquerque Sales, Tatiana de Arruda Campos Brasil de Souza, Diogo Borges Lima, Fabio C. Gozzo, Richard Hemmi Valente, Rosario Duran, Valmir C. Barbosa, and Paulo C. Carvalho*

Cite This: https://doi.org/10.1021/jasms.3c00063

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Complex protein mixtures typically generate many tandem mass spectra produced by different peptides coisolated in the gas phase. Widely adopted proteomic data analysis environments usually fail to identify most of these spectra, succeeding at best in identifying only one of the multiple cofragmenting peptides. We present PatternLab V (PLV), an updated version of PatternLab that integrates the YADA 3 deconvolution algorithm to handle such cases efficiently. In general, we expect an increase of 10% in spectral identifications when dealing with complex proteomic samples. PLV is freely available at http://patternlabforproteomics.org.



One of the aims of shotgun proteomics is to identify as many peptide ions as possible[1] by isolating them in the gas phase and obtaining their tandem mass spectra. The experimental spectra are then compared to those theoretically generated from a sequence database, and the final list of candidates is filtered according to specific criteria to obtain a 1% false-discovery rate. As it turns out, generally, the proteomic samples routinely analyzed are complex ones such as cell lysates, biopsies, or body fluids. The tryptic digest of these complex samples gives rise to even more complex mixtures consisting of millions of peptides that typically coelute during reversed-phase chromatography coupled with the mass spectrometer at hand. It is thus no coincidence that, during coelution, many of those peptides have very similar mass-to-charge ratios, commonly less than 1 m/z. Widely adopted mass spectrometers typically use quadrupoles or linear traps for precursor ion isolation, being therefore limited to isolating ions of approximately ±1 Da. Consequently, other peptides of close m/z manage to be coisolated and cofragmented with the precursor ion selected by the instrument.

It is well-known that multiplexed spectra can nowadays be handled efficiently. The first report points to Venable et al., who debuted Data-Independent Acquisition (DIA) by acquiring sequential isolation windows on an LTQ mass spectrometer.[2] At the time, the limitations inherent to their instruments capped both identifications and the broad applicability of their method. Still, unequivocally, they had succeeded in anticipating one of the next big things in proteomics. In what followed, Carvalho et al., working with an Orbitrap-XL, were the first to acquire high-resolution MS1 data together with the isolation windows in an approach named Extended Data-Independent Analysis (XDIA) that allowed the precursors in the isolation window to be known beforehand.[3] Later, Gillet et al. applied the same strategy to other TOF instruments and relied on spectral libraries for data analysis, calling the entire package SWATH.[4] More recently, we introduced an approach combining Data-Dependent Acquisition (DDA) with DIA and named it Mixed-Data Acquisition.[5]

Be that as it may, different groups have argued about the advantages and disadvantages of choosing DDA or DIA, and a point of view on this subject is beyond the scope of this note. One of the advantages of DDA is that it provides "cleaner" mass spectra that may be fitter for *de novo* sequencing and open searches[6] and even for handling more complex scenarios such as cross-linking mass spectrometry.[7] As mentioned, in DDA, some ions slip in together with the precursor and produce multiplexed spectra.

Received: February 17, 2023
Revised: March 7, 2023
Accepted: March 8, 2023

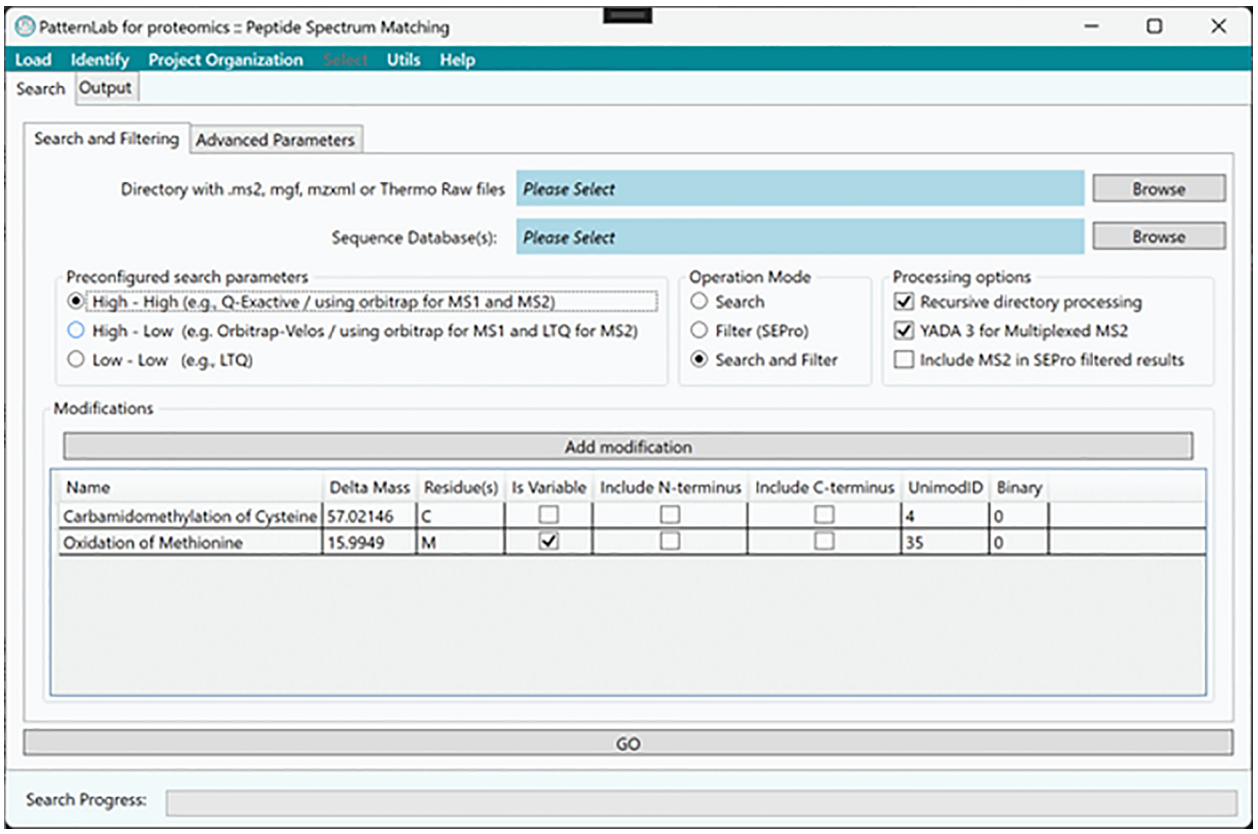


Figure 1. A screenshot of the PLV search-engine GUI. The 'YADA 3 for Multiplexed MS2' checkbox, located in the "Processing options" groupbox, is checked by default. This enables PLV to identify more than one peptide when dealing with multiplexed tandem mass spectra.

We recently introduced the third version of Yet Another Deconvolution Algorithm (YADA 3).[8] As we have shown since the first version appeared,[9] YADA 3 can be used coupled with a search engine to single out spectra resulting from two coisolated precursors. If done correctly, this can increase search engine results. We have also shown YADA 3 to be 1 order of magnitude faster than the previous fastest deconvolution algorithm and capable of listing more isotopic envelopes, with fewer errors, than its competitors (as estimated by our mass-defect approach).

YADA 3's exceptional speed and high confidence motivated us to integrate its algorithm natively into PatternLab V (PLV).[10] A checkbox has been made available in the search engine's Graphical User Interface (GUI) and will be checked by default (Figure 1). This will trigger YADA to automatically tell PLV which spectra are multiplexed so that the internal search engine can efficiently deal with them. In essence, with this upgrade, we are making it possible for the identification rate to be increased. Previously, we reported, for both the original YADA and YADA 3, that spectral identification had increased by approximately 10%. These results have now been repeated with PLV's updated version. We expect these gains to vary depending on sample complexity, chromatography, and experimental setup; for example, shorter chromatographies will generate more coeluting peptides, raising the chances of multiplexed samples.

We anticipate that this integration of YADA 3 into PLV will benefit PLV users immediately by increasing peptide identifications and protein coverage. PatternLab is freely available on the project's Web site, http://www.patternlabforproteomics.org.

AUTHOR INFORMATION

Corresponding Author

Paulo C. Carvalho − *Laboratory for Structural and Computational Proteomics, Carlos Chagas Institute, Fiocruz-Parana 81310-020, Brazil;* orcid.org/0000-0001-6530-3350; Email: paulo@pcarvalho.com

Authors

Milan A. Clasen − *Laboratory for Structural and Computational Proteomics, Carlos Chagas Institute, Fiocruz-Parana 81310-020, Brazil;* orcid.org/0000-0002-9118-7581

Marlon D. M. Santos − *Laboratory for Structural and Computational Proteomics, Carlos Chagas Institute, Fiocruz-Parana 81310-020, Brazil;* orcid.org/0000-0002-1178-1266

Louise Ulrich Kurt − *Laboratory for Structural and Computational Proteomics, Carlos Chagas Institute, Fiocruz-Parana 81310-020, Brazil;* orcid.org/0000-0002-9703-7226

Juliana Fischer − *Laboratory for Structural and Computational Proteomics, Carlos Chagas Institute, Fiocruz-Parana 81310-020, Brazil;* orcid.org/0000-0001-5529-4366

Amanda C. Camillo-Andrade − *Laboratory for Structural and Computational Proteomics, Carlos Chagas Institute, Fiocruz-Parana 81310-020, Brazil;* orcid.org/0000-0003-4642-3984

Lucas Albuquerque Sales − *Laboratory for Structural and Computational Proteomics, Carlos Chagas Institute, Fiocruz-Parana 81310-020, Brazil*

https://doi.org/10.1021/jasms.3c00063
J. Am. Soc. Mass Spectrom. XXXX, XXX, XXX−XXX

Tatiana de Arruda Campos Brasil de Souza − *Laboratory for Structural and Computational Proteomics, Carlos Chagas Institute, Fiocruz-Parana 81310-020, Brazil*

Diogo Borges Lima − *Department of Structural Biology, Leibniz—Forschungsinstitut für Molekulare Pharmakologie (FMP), Berlin 13125, Germany;* orcid.org/0000-0001-6056-0825

Fabio C. Gozzo − *Dalton Mass Spectrometry Laboratory, Unicamp, Campinas 13083-970, Brazil*

Richard Hemmi Valente − *Laboratory of Toxinology, Oswaldo Cruz Institute, Fiocruz, Rio de Janeiro 21040-900, Brazil;* orcid.org/0000-0002-6562-7571

Rosario Duran − *Analytical Biochemistry and Proteomics Unit, Institut Pasteur de Montevideo/IIBCE, Montevideo 11400, Uruguay*

Valmir C. Barbosa − *Systems Engineering and Computer Science Program, Federal University of Rio de Janeiro, Rio de Janeiro 21941-914, Brazil;* orcid.org/0000-0003-1010-7170

Complete contact information is available at: https://pubs.acs.org/10.1021/jasms.3c00063

Author Contributions

[+]M.A.C. and M.D.M.S. contributed equally.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work received partial financial support from CNPq, Inova Fiocruz, Fiocruz PEP, Fundação Araucária, CAPES, and a FAPERJ BBP grant. The authors acknowledge support from the Mass Spectrometry Facility RPT02H at Fiocruz-Paraná.

■ REFERENCES

(1) Eng, J. K.; McCormack, A. L.; Yates, J. R. An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *J. Am. Soc. Mass Spectrom.* **1994**, *5* (11), 976−989.

(2) Venable, J. D.; Dong, M.-Q.; Wohlschlegel, J.; Dillin, A.; Yates, J. R. Automated Approach for Quantitative Analysis of Complex Peptide Mixtures from Tandem Mass Spectra. *Nat. Methods* **2004**, *1* (1), 39−45.

(3) Carvalho, P. C.; Han, X.; Xu, T.; Cociorva, D.; Carvalho, M. da G.; Barbosa, V. C.; Yates, J. R. XDIA: Improving on the Label-Free Data-Independent Analysis. *Bioinformatics* **2010**, *26* (6), 847−848.

(4) Gillet, L. C.; Navarro, P.; Tate, S.; Röst, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. Targeted Data Extraction of the MS/MS Spectra Generated by Data-Independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Mol. Cell Proteomics* **2012**, *11* (6), O111.016717.

(5) Santos, M. D. M.; Camillo-Andrade, A. C.; Kurt, L. U.; Clasen, M. A.; Lyra, E.; Gozzo, F. C.; Batista, M.; Valente, R. H.; Brunoro, G. V. F.; Barbosa, V. C.; Fischer, J. S. G.; Carvalho, P. C. Mixed-Data Acquisition: Next-Generation Quantitative Proteomics Data Acquisition. *J. Proteomics* **2020**, *222*, 103803.

(6) Kong, A. T.; Leprevost, F. V.; Avtonomov, D. M.; Mellacheruvu, D.; Nesvizhskii, A. I. MSFragger: Ultrafast and Comprehensive Peptide Identification in Mass Spectrometry−Based Proteomics. *Nat. Methods* **2017**, *14* (5), 513−520.

(7) Lima, D. B.; Melchior, J. T.; Morris, J.; Barbosa, V. C.; Chamot-Rooke, J.; Fioramonte, M.; Souza, T. A. C. B.; Fischer, J. S. G.; Gozzo, F. C.; Carvalho, P. C.; Davidson, W. S. Characterization of Homodimer Interfaces with Cross-Linking Mass Spectrometry and Isotopically Labeled Proteins. *Nat. Protoc.* **2018**, *13* (3), 431−458.

(8) Clasen, M. A.; Kurt, L. U.; Santos, M. D. M.; Lima, D. B.; Liu, F.; Gozzo, F. C.; Barbosa, V. C.; Carvalho, P. C. Increasing Confidence in Proteomic Spectral Deconvolution through Mass Defect. *Bioinformatics* **2022**, *38* (22), 5119−5120.

(9) Carvalho, P. C.; Xu, T.; Han, X.; Cociorva, D.; Barbosa, V. C.; Yates, J. R. YADA: A Tool for Taking the Most out of High-Resolution Spectra. *Bioinformatics* **2009**, *25* (20), 2734−2736.

(10) Santos, M. D. M.; Lima, D. B.; Fischer, J. S. G.; Clasen, M. A.; Kurt, L. U.; Camillo-Andrade, A. C.; Monteiro, L. C.; de Aquino, P. F.; Neves-Ferreira, A. G. C.; Valente, R. H.; Trugilho, M. R. O.; Brunoro, G. V. F.; Souza, T. A. C. B.; Santos, R. M.; Batista, M.; Gozzo, F. C.; Durán, R.; Yates, J. R.; Barbosa, V. C.; Carvalho, P. C. Simple, Efficient and Thorough Shotgun Proteomic Analysis with PatternLab V. *Nat. Protoc.* **2022**, *17*, 1553−1578.