HAL open science

TDFragMapper: a visualization tool for evaluating experimental parameters in top-down proteomics

Jonathan Dhenin, Diogo B Lima, Mathieu Dupré, Julia Chamot-Rooke

To cite this version:

Jonathan Dhenin, Diogo B Lima, Mathieu Dupré, Julia Chamot-Rooke. TDFragMapper: a visualization tool for evaluating experimental parameters in top-down proteomics. Bioinformatics, 2021, pp.btab784. 10.1093/bioinformatics/btab784 . pasteur-03441695

HAL Id: pasteur-03441695

https://pasteur.hal.science/pasteur-03441695v1

Submitted on 22 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Sequence analysis

TDFragMapper: a visualization tool for evaluating experimental parameters in top-down proteomics

Jonathan Dhenin[1,2,†], Diogo B Lima[3,†,*], Mathieu Dupré[1], Julia Chamot-Rooke[1,*]

[1]Mass Spectrometry for Biology Unit, CNRS, USR2000, Institut Pasteur, Paris, France, [2]Université de Paris, Sorbonne Paris Cité, Paris, France, [3]Leibniz-Forschungsinstitut für Molekulare Pharmakologie, Berlin, Germany

*To whom correspondence should be addressed.

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

Associate Editor: XXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Motivation: We present a new software-tool allowing an easy visualization of fragment ions and thus a rapid evaluation of key experimental parameters on the sequence coverage obtained for the MS/MS analysis of intact proteins. Our tool can process data obtained from various deconvolution and fragment assignment software.

Results: We demonstrate that TDFragMapper can rapidly highlight the experimental fragmentation parameters that are critical to the characterization of intact proteins of various size using top-down proteomics.

Availability: TDFragMapper, a demonstration video and user tutorial are freely available for academic use at https://msbio.pasteur.fr/tdfragmapper; all data are thus available from the ProteomeXchange consortium (identifier PXD024643).

Contact: diogobor@gmail.com or julia.chamot-rooke@pasteur.fr

Supplementary information: Supplementary data are available at *Bioinformatics* online

1 Introduction

Top-down proteomics (TDP) is a powerful technology allowing the characterization of proteins at the proteoform level using high-resolution tandem mass spectrometry (MS/MS). Proteoforms correspond to the different forms of a protein arising from all combinatorial sources of variation from a single gene including combinations of genetic variation, alternative splicing, and post-translational modifications (PTM)(Smith and Kelleher, 2018). The complete characterization of proteoforms often requires the use of several complementary fragmentation techniques, such as collision-induced dissociation (CID), electron transfer dissociation (ETD) or ultraviolet photodissociation (UVPD)(Fornelli *et al.*, 2018). In contrast to bottom-up proteomics, the experimental parameters used for the fragmentation in TDP, such as the activation energy or the charge state of the precursor ion, can significantly affect the quality of MS/MS data and therefore the protein sequence coverage (Brunner *et al.*, 2015). Although existing tools are capable of matching a list of fragment ions to a protein sequence (ProSight Lite or MASH Explorer for instance)(Fellers *et al.*, 2015; Wu *et al.*, 2020), there is currently no computational tool allowing to visualize fragments arising from diverse MS/MS experiments on a unique fragmentation map without losing information on the contribution and the specificity of each experiment. Current tools provide a unique fragmentation map per MS/MS experiment, and thus their comparison, in particular when multiple parameters are assessed, is both difficult and time-consuming. Moreover, although the intensity of fragment ions can represent a precious source of information when interpretating MS/MS data, it is often absent of fragmentation maps. To circumvent these limitations, we introduce TDFragMapper, a novel software-tool that can display and combine pre-assigned fragment ions achieved from various MS/MS experiments on a unique protein sequence, keeping an easy access to the individual contribution of each experiment and to the intensity of deconvoluted fragment ions. Our tool makes it possible to rapidly compare experimental parameters such as the type of fragmentation, the activation level or the precursor charge state in the MS/MS analysis of intact proteins. In what follows, we use TDFragMapper to evaluate the influence of essential experimental parameters in the TDP analysis of a standard mixture of intact proteins.

2 Material and Methods

We analyzed a standard mixture of six proteins (Thermo Scientific Pierce Intact Protein Standard Mix) in LC-MS/MS on an Orbitrap Fusion™ Lumos™ mass spectrometer using multiple experimental conditions, as described in the Supplementary Material. Raw files were deconvoluted using FreeStyle™ (v. 1.6.75.20) and the lists of deconvoluted ion masses were imported into ProSight Lite (v. 1.4) in order to assign fragments. Lists of assigned fragments were then exported and used with deconvoluted data and the protein sequence as input for TDFragMapper. The software was programmed in C# with .NET Framework 4.8 and requires a computer with Windows 10 or later, and at least 8 GB of RAM. To evaluate the individual contributions of key-parameters in TDP experiments, and rapidly assess the combination leading to the best sequence coverage, the software is organized in three major interfaces: an input interface, a filter interface, and a graphical visualization module.

© The Author(s) 2021. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Downloaded from https://academic.oup.com/bioinformatics/advance-article/doi/10.1093/bioinformatics/btab784/6430971 by Institut Pasteur user on 23 November 2021

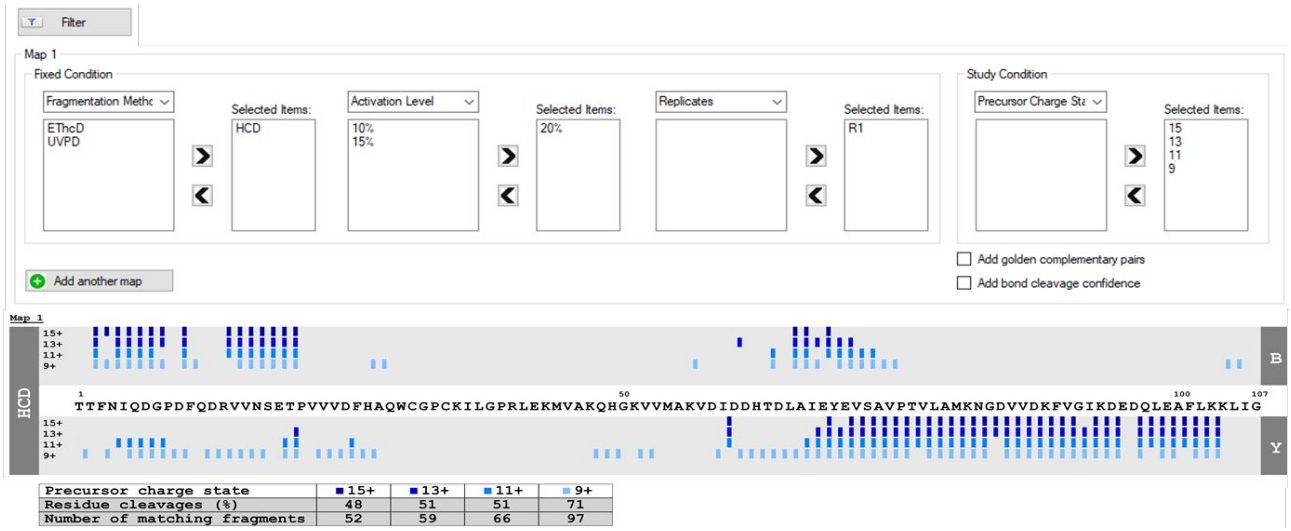

Figure 1: Screenshot of the filter interface of TDFragMapper (top panel) and the visualization module with the corresponding fragmentation map of human Thioredoxin (11 kDa) obtained when varying the precursor charge state in HCD with NCE 20% (bottom panel). Fragmentation maps obtained for larger proteins (21 and 29 kDa) are depicted in Figures S6 and S7.

Besides data files, the input interface allows the user to upload the putative protein sequence as a *.fasta file, any identified PTM and the experimental parameters used for each fragmentation experiment in an easy-to-fill table format. Two template files are also accessible in the help section to process results achieved with any deconvolution and fragment assignment tools. A nomenclature is proposed to ensure a standardized filling of these experimental parameters. The drop-down menus and the arrow buttons of the filter interface allow the user to quickly set some experimental parameters and study their influence one by one (Figure 1, upper panel). Once filtered, the fragments are displayed in a visualization module and mapped onto the protein sequence (Figure 1, lower panel). Fragments are represented by vertical bars arranged in rows and colored according to the MS/MS experiment from which the fragment arises. N-terminal fragments (a-, b- and c- ions) and C-terminal fragments (x-, y- and z-ions) are displayed respectively above and under the protein sequence. PTMs are depicted with an orange dot above the modified amino acid. A legend table is associated to each fragmentation map with the color code of fragments, the percentage of residue cleavages and the number of matching fragments for each experimental parameter under study.

3 Results & Discussion

Displaying fragments obtained from different MS/MS experiments onto a unique linear protein sequence with a specific color code makes it possible to highlight at a glance the contribution of each experimental parameter. TDFragMapper can be used for example to spot the dependency of a protein sequence coverage on the selected precursor charge state when using a particular fragmentation method (Figure 1, lower panel) or to assess the reproducibility of a particular fragmentation method across technical replicates. For instance, in Figure 1, it is easy to visualize that the charge state leading to the best sequence coverage for the HCD fragmentation of thioredoxin is the +9 one. Moreover, the user can add extra layer of information to the fragmentation map, such as the position and the number of golden complementary pairs (Kelleher *et al.*, 1999; Horn *et al.*, 2000)(Figures S1 and S2) or the intensity of deconvoluted fragment ions. This intensity option is easily accessible from the graphical visualization module and can be used to compare the abundance of common fragments across multiple MS/MS experiments on a single map (Figure S3). The intensity option can be used to improve confidence in the localization of a PTM as shown in Figure S4. Once identified by the user thanks to the different fragmentation maps and the information provided by the previously described features, the results of the best fragmentation experiments can be summed using the merging option onto a single fragmentation map. This final map displays only the N-terminal and C-terminal cleavages regardless of the ion type and a final residue cleavage is computed, as can be seen in Figure S5.

Fragment maps can be exported as *.tiff, *.png or *.jpg image files. TDFragMapper also allows to export a summary report as PDF® file containing the information of uploaded data files and all the parameters used to create the maps. Finally, the work session and the results can be directly saved in an owner format (*.tdfm) that can be retrieved and used later. TDFragMapper is highly complementary to other existing software tools allowing the results of targeted TDP experiments to be visualized on a single fragmentation map.

4 Final Remarks

We anticipate that TDFragMapper will ease the selection of optimal fragmentation parameters in order to increase the confidence in proteoform characterization in TDP experiments, including the precise localization of post-translational modifications. To facilitate its use, a tutorial with all functionalities is included in the tool's website (Borges Lima *et al.*, 2021) and can also be accessed through the help menu.

Funding

This work has received funding from the European Union's Horizon 2020 research and innovation program under grant agreements 829157 and 823839.

Conflict of Interest: none declared.

References

Borges Lima,D. *et al.* (2021) Using TDFragMapper: a high-performance tool to assess experimental parameters in top-down proteomics.

Brunner,A.M. *et al.* (2015) Benchmarking Multiple Fragmentation Methods on an

Downloaded from https://academic.oup.com/bioinformatics/advance-article/doi/10.1093/bioinformatics/btab784/6430971 by Institut Pasteur user on 23 November 2021

Orbitrap Fusion for Top-down Phospho-Proteoform Characterization. *Anal. Chem.*, **87**, 4152–4158.

Fellers,R.T. *et al.* (2015) ProSight Lite: Graphical software to analyze top-down mass spectrometry data. *PROTEOMICS*, **15**, 1235–1238.

Fornelli,L. *et al.* (2018) Accurate Sequence Analysis of a Monoclonal Antibody by Top-Down and Middle-Down Orbitrap Mass Spectrometry Applying Multiple Ion Activation Techniques. *Analytical Chemistry*, **90**, 8421–8429.

Horn,D.M. *et al.* (2000) Automated de novo sequencing of proteins by tandem high-resolution mass spectrometry. *PNAS*, **97**, 10313–10317.

Kelleher,N.L. *et al.* (1999) Top Down versus Bottom Up Protein Characterization by Tandem High-Resolution Mass Spectrometry. *J. Am. Chem. Soc.*, **121**, 806–812.

Smith,L.M. and Kelleher,N.L. (2018) Proteoforms as the next proteomics currency. *Science*, **359**, 1106–1107.

Wu,Z. *et al.* (2020) MASH Explorer: A Universal Software Environment for Top-Down Proteomics. *J Proteome Res*, **19**, 3867–3876.

Downloaded from https://academic.oup.com/bioinformatics/advance-article/doi/10.1093/bioinformatics/btab784/6430971 by Institut Pasteur user on 23 November 2021