Supplementary Material

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

Jonathan Dhenin¹,²,†, Diogo B Lima³,⁴, Mathieu Dupré¹, Julia Chamot-Rooke¹,*

¹Mass Spectrometry for Biology Unit, CNRS USR 2000, Institut Pasteur, Paris, France, ²Université de Paris, Sorbonne Paris Cité, Paris, France, ³Leibniz-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

Materials & Methods

Sample preparation

The Thermo Scientific™ Pierce™ Intact Protein standard mix is composed of the six following intact proteins: human IGF-I LR3 (9,111.47 Da), human Thioredoxin (11,865.52 Da), Streptococcus dysgalactiae Protein G (21,442.61 Da), Bovine Carbonic Anhydrase II (28,981.29), Streptococcus Protein AG (50,459.74 Da) and Escherichia coli Exo Klenow (68,001.15 Da). 76 µg of lyophilized protein mixture were reconstituted in 380 µL of solvent A (98% H2O, 2% ACN, 0.1% FA) to a final concentration of 0.2 µg/µL.

LC-MS/MS

A Dionex UltiMate 3000 RSLC Nano System coupled to an Orbitrap Fusion™ Lumos™ mass spectrometer fitted with a nano-electrospray ionization source (Thermo-Scientific) was used. Five µL of protein sample were loaded at a flow rate of 10 µL.min⁻¹ onto an in-house packed C4 (5µm, Reprosil) trap column (0.150 mm i.d. x 30 mm) and separated at a flow rate of 1 µL.min⁻¹ using a C4 (5 µm, Reprosil) column (0.150 mm i.d. x 400 mm). The following gradient was used: 2% solvent B (20% H2O, 80% ACN, 0.1% FA) from 0-5 min; 20% B at 6 min.; 35% B at 7 min.; 60% B at 16 min.; 99% B from 17-20 min.; and 2% B from 20.2-40 min.

A first LC-MS experiment was acquired at 15,000 resolving power (at m/z 400) with a scan range set to 550-2,000 m/z, 5 microscans (µscans) per MS scan, an automatic gain control (AGC) target value of 5x10⁵ and maximum injection time of 50 ms. Fragmentation data were recorded using targeted LC-MS/MS experiments. Four precursor charge states were chosen for each protein across their respective charge state distribution and isolated by the quadrupole and subjected to fragmentation with a maximum of two charge states per chromatographic run for each protein (Table S1). MS/MS scans were acquired at 120,000 resolving power (at m/z 400) with an isolation width of 1.6 m/z, 5 µscans, an AGC target value of 5x10⁵ and maximum injection time of 246 ms. Higher-energy collisional dissociation with NCE of 10, 15 and 20% (HCD), electron transfer dissociation with 5, 10 and 15 ms of reaction time and a supplemental higher-energy collisional dissociation with NCE of respectively 5, 10 and 15% (ETThcD), ultraviolet photodissociation at 213 nm with 40, 50 and 60 ms of reaction time (UVPD) were used for the fragmentation of intact proteins.

Table S1. Targeted m/z and the corresponding charge state for each protein

| Protein          | Targeted m/z | Charge state | Protein          | Targeted m/z | Charge state |
|------------------|--------------|--------------|------------------|--------------|--------------|
| IGF-I LR3        | 916.965      | 10+          | BCA2             | 763.697      | 38+          |
|                  | 1013.209     | 9+           |                  | 829.008      | 35+          |
|                  | 1139.704     | 8+           |                  | 906.658      | 32+          |
|                  | 1302.329     | 7+           |                  | 1000.357     | 29+          |
| Thioredoxin      | 791.973      | 15+          |                  | 935.435      | 54+          |
|                  | 913.602      | 13+          | Protein AG       | 1010.179     | 50+          |
|                  | 1079.500     | 11+          |                  | 1097.943     | 46+          |
|                  | 1319.240     | 9+           |                  | 1202.416     | 42+          |
| Protein G        | 858.691      | 25+          | Exo Klenow       | 756.575      | 90+          |
|                  | 975.592      | 22+          |                  | 801.01       | 85+          |
|                  | 1129.466     | 19+          |                  | 851.02       | 80+          |
|                  | 1341.129     | 16+          |                  | 907.69       | 75+          |
Data analysis

MS/MS spectral data were first deconvoluted and deisotoped using the Xtract algorithm embedded in FreeStyle™ v1.6.75.20 (Thermo-Scientific) using a fit factor of 80%, a remainder threshold of 25% and a S/N threshold of 3. Scans were averaged across the width of the chromatographic peak. Deconvoluted ion masses were then exported as *.xls files and uploaded into ProSight Lite v1.4 (Fellers et al., 2015) with the appropriate protein sequence. Assigned fragments were finally exported as *.xlsx files. Both *.xls and *.xlsx files were used as input data for TDFragMapper.

Results

The addition of golden complementary pairs onto the fragmentation map, as displayed in Figure S1, is generated instantly when selecting the corresponding option in the filter interface of TDFragMapper. This option makes it easier to the user to localize quickly the golden complementary pairs. A golden complementary pair is a pair of fragment ions (a/x, b/y or c/z) that have been formed by cleavage between the same pair of amino acids (Horn et al., 2000; Kelleher et al., 1999). Since the sum of the masses of the two fragments equals the mass of the targeted protein, golden complementary pairs can greatly enhance confidence in protein characterization (Figure S2).

Figure S1: Fragmentation map of human Thioredoxin obtained when varying the precursor charge state in HCD with NCE 20%, with the “Golden complementary pairs” option set on.

Figure S2: Golden complementary pairs generated in tandem mass spectrometry of intact proteins

The overlay of the intensity of deconvoluted fragments with their localization, as displayed in Figure S3, is generated instantly by the “Intensity” option embedded in TDFragMapper. This option is of great advantage to quickly spot the most intense fragments for each condition represented on the map. Moreover, variations in the abundance of fragments can be easily spotted. In Figure S3, it can be easily seen that the y-fragment between Val74 and Pro75 is the most intense fragment regardless of the selected precursor charge state.

Figure S3: Fragmentation map of human Thioredoxin obtained when varying the precursor charge state in HCD with NCE 20%, with the “Intensity” option set on.
The intensity option can also be used to improve confidence in the localization of a post-translational modification. As an example, we used the top-down dataset of a recently published article on *de novo* sequencing of antibody light chain proteoforms (Dupré *et al.*, 2021) in which one proteoform was found to be modified with an O-glycan HexNAc(1)dHex(1) on either Ser160, Ser163 or Ser169. HCD fragmentation maps of the three possible proteoforms were generated with TDFragMapper and compared. The visualization of these data (in particular the exclusive fragment ions and their intensity) indicates that there is probably a mixture of proteoforms (with a preference on Ser160) and not a single modification site (*Figure S4*).

*Figure S4: HCD fragmentation maps of an antibody light chain with 3 potential sites of O-glycosylation, Ser160 (top), Ser163 (middle) and Ser169 (bottom)*

The complete characterization of proteins by top-down mass spectrometry often requires the use of diverse fragmentation techniques (Fornelli *et al.*, 2018; Brunner *et al.*, 2015). The merging option can be used to sum all the fragments arising from the best fragmentation experiments. A single fragmentation map is created and gathers all a-, b- and c-ions into N-terminal cleavages and all x-, y- and z-ions into C-terminal cleavages. In *Figure S5*, the fragmentation results obtained with HCD (NCE 20%) and with EThcD (5 ms + NCE 5%) are displayed in separate maps. The best fragmentation results, here the fragments obtained with the 9+ precursor in HCD and the 15+ precursor in EThcD, are merged into the final fragmentation map displayed at the bottom with a final residue cleavage of 92%.
Figure S5: Fragmentation maps of human Thioredoxin obtained when varying the precursor charge state in EThcD with 5ms + NCE 5% (top), in HCD with NCE 20% (middle) and with the best results of EThcD and HCD combined using the “Merging” option (bottom).

TDFragmenter is also of advantage for the analysis of larger proteins. Fragmentation parameters of both Streptococcus dysgalactiae Protein G (21 kDa) and bovine carbonic anhydrase II (29 kDa) have been easily optimized to reach respectively 63% and 50% residue cleavages in only 2 fragmentation experiments (Figures S6 and S7). The visualization provided by TDFragmenter highlights the different dependency of HCD fragmentation on the precursor charge state for 2 proteins of close molecular weight.
Figure S6: Fragmentation maps of *Streptococcus dysgalactiae* Protein G obtained when varying the precursor charge state in HCD 20% (top), in EThcD with 5ms + NCE 5% (middle) and with the best results of EThcD (25+) and HCD (16+) combined using the “Merging” option (bottom). The final residue cleavage percentage reaches 63%.
Figure S7: Fragmentation maps of bovine carbonic anhydrase II obtained when varying the precursor charge state in HCD 20 % (top), in EThcD with 5ms + NCE 5% (middle) and with the best results of EThcD (35+) and HCD (32+) combined using the “Merging” option (bottom). The final residue cleavage percentage reaches 50%.
References
Brunner, A.M. et al. (2015) Benchmarking Multiple Fragmentation Methods on an Orbitrap Fusion for Top-down Phospho-Proteoform Characterization. *Anal. Chem.*, **87**, 4152–4158.
Dupré, M. et al. (2021) De Novo Sequencing of Antibody Light Chain Proteoforms from Patients with Multiple Myeloma. *Anal. Chem.*, **93**, 10627–10634.
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. *Anal. Chem.*, **90**, 8421–8429.
Horn, D.M. et al. (2000) Automated de novo sequencing of proteins by tandem high-resolution mass spectrometry. *Proc. Natl. Acad. Sci.*, **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.
Wu, Z. et al. (2020) MASH Explorer: A Universal Software Environment for Top-Down Proteomics. *J. Proteome Res.*, **19**, 3867–3876.