Published in final edited form as:

Methods Mol Biol. 2017 ; 1558: 381–394. doi:10.1007/978-1-4939-6783-4_18.

Bioinformatics Analysis of Top-Down Mass Spectrometry Data with ProSight Lite

Caroline J. DeHart, Ryan T. Fellers, Luca Fornelli, Neil L. Kelleher, and Paul M. Thomas

Abstract

Traditional bottom-up mass spectrometry-based proteomics relies on the use of an enzyme, often trypsin, to generate small peptides (typically < 25 amino acids long). In top-down proteomics, proteins remain intact and are directly measured within the mass spectrometer. This technique, while inherently simpler than bottom-up proteomics, generates data which must be processed and analyzed using software tools “purpose-built” for the job. In this chapter, we will show the analysis of intact protein spectra through deconvolution, deisotoping, and searching with ProSight Lite, a free, vendor-agnostic tool for the analysis of top-down mass spectrometry data. We will illustrate with two examples of intact protein fragmentation spectra and discuss the iterative use of the software to characterize proteoforms and discover the sites of post-translational modifications.

Keywords

Top-down; Mass spectrometry; Proteomics; ProSight Lite; Intact protein; Bioinformatics

1 Introduction

Complementing the speed and sensitivity of bottom-up proteomics [1, 2], top-down proteomics [3] offers a comprehensive view of proteoforms present in the sample (for a more detailed description of a proteoform, see Note 1) [4]. It is important to distinguish between top-down mass spectrometry and top-down proteomics. In top-down mass spectrometry, one or a few proteoforms are isolated and studied. This technique is typified by studies of histones [5–7] and immunoglobulins [8–10], among others, including the routine analysis of small (< 40 kDa) proteins. In top-down proteomics, a whole, unknown proteome is typically isolated, fractionated, and analyzed by mass spectrometry [11–13]. The tools used for the study of top-down mass spectrometry and top-down proteomics are related, but fundamentally different. Top-down proteomics requires searching tandem MS data against a database of known proteins and proteoforms [14, 15], while top-down mass spectrometry is simpler. In the latter case, the experimenter generally knows the protein under study, but may want to characterize the location of post-translational modifications [5] or to better understand a new protein fragmentation technique [16]. Here we illustrate, with real data use cases, how to deconvolute and deisotope top-down mass spectrometry data with

¹A proteoform is a single instance of a group of protein products arising from a specific gene incorporating multiple sources of potential variation including alternative splicing, single-nucleotide polymorphism, and post-translational modifications, among others [4].

QualBrowser and Xtract (ThermoFisher, San Jose, CA) followed by analysis with ProSight Lite [17], a free tool for intact protein data analysis.

2 Materials

1. ProSight Lite [17] downloaded from <http://prosightlite.northwestern.edu>. This must be installed on a Microsoft Windows computer with .NET Framework 4.5.1 or greater.
2. Top-down mass spectrometry data. A ThermoFisher .raw file for a top-down standard used in this study has been provided at <http://hdl.handle.net/2022/20816>. In addition to the raw data, the output from Subheading 3.1 (below) is provided as PostXtract .raw files and text files containing the deconvoluted and deisotoped masses for use as input in Subheadings 3.2 and 3.3. A brief description of how the sample was prepared is found in Note 2.
3. Internet access to <http://www.uniprot.org>. This is required to automatically extract protein sequence information from a UniProt [18] entry.
4. (Optional). Software tools to view, deconvolute, and deisotope mass spectra. Many vendors supply a proprietary algorithm with their instruments. In this method, Xtract was used within QualBrowser from ThermoFisher Scientific. Other free options for the deconvolution and deisotoping of top-down mass spectrometry data include MS-Deconv [19] (available at <http://bix.ucsd.edu/projects/msdeconv/>) or YADA [20] (available at <http://fields.scripps.edu/yada/>); however, a full discussion of their use is beyond the scope of this chapter.
5. (Optional). Software tools to convert mass spectrometry data into an extensible markup language format such as .mzML and .mzXML. This may be required by deconvolution/deisotoping tools. One suggestion would be ProteoWizard [21], found at <http://proteowizard.sourceforge.net/downloads.shtml>.

3 Methods

3.1 Deconvolution and Deisotoping of Raw Data

1. For this exercise, download **TopDownStandard.raw** from the repository and open the file in QualBrowser. We will deconvolute and deisotope data for two of the five proteins present in the file, ubiquitin and carbonic anhydrase, as shown in Fig. 1 (see Note 2 for more information). The goal of the steps to follow is to generate zero-charge, monoisotopic mass spectra for the precursor and fragmentation scans. Instructions presented below will work with ThermoFisher .raw files. Other workflows can be generated with other vendor-

²The top-down standard is a solution containing bovine carbonic anhydrase (Sigma C2624), equine myoglobin (Sigma M5696), bovine trypsinogen (Sigma T1143), and bovine ubiquitin (Sigma U6253). A nanoLCMS run is performed in which a sample containing 0.6 pmol carbonic anhydrase, 1.1 pmol myoglobin, 0.5 pmol trypsinogen, and 0.1 pmol ubiquitin is injected onto a 75 μ m ID $\times$ 15 cm PLRP-S column (5 μ m d_p, Agilent) and eluted from 5 % acetonitrile in water + 0.1 % formic acid to 52 % acetonitrile in water + 0.1 % formic acid over 27 min at a flowrate of 300 nL/min. Data are acquired in data-dependent MS/MS mode on a ThermoFisher Velos Orbitrap Elite targeting the top 1 *m/z* value. That species is isolated in a 15 *m/z* window and HCD fragmentation occurs at 25 % NCE. Bovine superoxide dismutase (SOD1) is present as a contaminant in bovine carbonic anhydrase.

provided software or using openly available tools as described in **Materials**. All output files from the steps below are published with the dataset as indicated above in Subheading 2, **item 2**.

2. To prepare data for ProSight Lite, we require the intact mass (MS1 data) for each protein and a list of MS2 fragments observed. In QualBrowser, display the intact mass spectrum for ubiquitin from scan 334 (for more in-depth explanations of QualBrowser commands, see <http://planetorbitrap.com/download.php?filename=524212aa70353.pdf>).
3. Display the Xtract dialog for scan 334 by pinning the mass spectrum tab. Right click to bring up the options menu, select the **Export** menu, and then click **Xtract**. If **Xtract** is not available, see Note 3.
4. Set the Xtract parameters as displayed in Fig. 2. For top-down mass spectrometry, monoisotopic, zero-charge data are typically required. Further explanation of these parameters may be found in Note 4. Select Xtract to generate a deisotoped, deconvoluted spectrum from scan 334.
5. Select the new window that appears by pinning it. Navigate to scan number two of the PostXtract scan. This scan contains the monoisotopic, zero-charge ($^{13}\text{C}_0$) mass spectrum. Scan 1 contains the zero-charge mass spectrum that has not been deisotoped (and thus contains all of the isotopes). Scan 3 contains the m/z values from the input spectrum that were deconvoluted and deisotoped by the Xtract algorithm. Scan 4 contains the remaining m/z values from the input spectrum that were *not* deconvoluted and deisotoped by the Xtract algorithm.

³If Xtract is not available within your copy of QualBrowser, contact your local ThermoFisher mass spectrometry sales representative to find out how to enable this feature.

⁴Description of Xtract parameters with QualBrowser:

- a. **Generate Masses Mode:** Specifies whether Xtract returns zero-charge (**M**) or protonated, singly charged (**M+**) data. In this chapter, we use **M** as the convention as it is more intuitive.
- b. **Mass Range:** Specifies the m/z range to be deconvoluted and deisotoped. This is typically set to the start m/z and end m/z of the scan being processed.
- c. **Resolution@400:** Specifies the mass resolving power (m/m) of the measured scan. This parameter is only used as a default if the resolving power cannot be determined from the data.
- d. **S/N Threshold:** Specifies the lowest signal-to-noise ratio that the Xtract algorithm will confer with trying to deconvolute and deisotope the mass spectrum. Altering the threshold lower will generate more masses; however, many of these masses arise from noise and may be incorrect. A higher S/N threshold will generate fewer noise-related masses, but also fewer real masses.
- e. **Fit Factor:** Specifies the minimum fit parameter used by the Xtract algorithm. Increasing this parameter requires a more perfect fit of the data to the theoretical spectrum to be considered real; reducing this parameter requires a poorer fit to be considered real.
- f. **Remainder:** Specifies the remainder of the fit that is left in the scan during Xtract analysis. This remainder is used to deconvolute overlapping, lower-intensity ion signals.
- g. **Average Table:** Specifies the model for the isotopic abundances to be used by Xtract. This should not be changed during typical operation.
- h. **Max Charge:** Specifies the maximum charge to be considered by the Xtract algorithm. This should be set at or above the highest charge expected in the mass spectrum.

6. Copy the monoisotopic, zero-charge ($^{13}\text{C}_0$) mass spectrum to the clipboard. Right click to bring up the options menu, select the **Export** menu, and then click **Clipboard (Exact Mass)** as shown in Fig. 3.
7. Open a spreadsheet program such as Microsoft Excel and paste the data from the clipboard. Note the most intense intact mass: **8,559.6556 Da**. Save the file as **UbiqMS1.xlsx**.
8. Repeat steps 3–7 with scan 330. Scan 330 is an HCD fragmentation scan of the 11+ charge state of ubiquitin. Save the resulting file as **UbiqMS2.xlsx**.
9. Repeat steps 3–7 with scan 556. Scan 556 is an intact mass spectrum of bovine carbonic anhydrase. Note the most intense intact mass: **29,006.8254 Da**. Save the resulting file as **CAMS1.xlsx**.
10. Repeat steps 3–7 with scan 561. Scan 561 is an HCD fragmentation scan of the 32+ charge state of carbonic anhydrase. Save the resulting file as **CAMS2.xlsx**.

3.2 ProSight Lite Analysis of Bovine Ubiquitin (Simple Analysis Case)

1. Open ProSight Lite and note the two pieces of information required: experimental data (MS1 and MS2) and sequence information (Fig. 4).
2. Click **Add Experimental Data**. Note the new window that appears (Fig. 5).
3. In **Precursor**, enter **8,559.6556** as determined in Subheading 3.1, **step 7**, above or copy from **UbiqMS1.xlsx**.
4. Open **UbiqMS2.xlsx** and copy the mass data from Column A. Paste these values into the **Fragments (one per line)** text box.
5. Select the following parameters (*see* Note 5 for more detailed descriptions), then click **Save**:
 - a. Precursor Mass Type: **Monoisotopic**
 - b. Mass Mode: **M (neutral)**
 - c. Fragmentation Method: **HCD**
 - d. Fragmentation Tolerance: **10 ppm**
6. Click on **Add Candidate Sequence**. Note that a new window appears (Fig. 6).

⁵Description of the experimental data parameters for ProSight Lite.

- a. **Precursor Mass Type** is derived from the Xtract results. Selecting scan 2 in Subheading 3.1, **step 5**, generates monoisotopic masses.
- b. **Mass Mode** is derived from the Xtract parameters; *see* Note 4.
- c. **Fragmentation Method** is determined by the fragmentation used in the mass spectrometer instrument method.
- d. **Fragmentation Tolerance** is intrinsic to your instrument's mass measurement accuracy. Setting this parameter too narrow will discard many real masses; however, setting this parameter too wide will negatively affect the confidence in the results [23].

7. In the sequence text box, enter the sequence of bovine ubiquitin:
MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIP
PDQQRLLFAGKQLEDGRTLSDYNIQKESTLHLVL RLRGG
8. Select **No Modification** for cysteine and methionine, and then click **Save**.
9. Results are obtained (Fig. 7) indicating a strong agreement between the experimental data and the candidate sequence. The P-score is 2.5×10^{-68} , multiple fragments cover both termini, and the intact mass differs from the theoretical by 0.04 Da or < 5 ppm. Flags pointing to the upper left indicate N-terminal fragments, while flags pointing to the lower right indicate C-terminal fragments. The fragment color is determined by the Roepstorff fragment type [22]: a/x are green, b/y are blue, and c/z are red. Scores are further described in Note 6.
10. Exports:
 - a. A list of matching fragment ions and their mass errors can be downloaded by selecting the Microsoft Excel icon to the upper right of the matching fragment table.
 - b. Fragmentation maps may be downloaded in PNG (raster graphic) or SVG (vector graphic) under the **Export** menu.
 - c. Results may be saved for future use or shared with collaborators using the PCML format under the **File... Save** menu.

3.3 ProSight Lite Analysis of Bovine Carbonic Anhydrase II (Iterative Analysis Case)

1. Open a new instance of ProSight Lite and **Add Experimental Data** as in Subheading 3.2, **step 2–5**, above with the following changes:
 - a. Use data from **CAMS1.xlsx** in Subheading 3.2, **step 3** (either enter **29,006.8254** as determined in Subheading 3.1, **step 9**, above or copy from the file).
 - b. Use data from **CAMS2.xlsx** in Subheading 3.2, **step 4**.
2. Click on **Add Candidate Sequence**. In the UniProt Accession text box (Fig. 6), enter **P00921**, and click **Search**. This populates the sequence text box with the amino acid sequence for bovine carbonic anhydrase II. Select **No Modification**

⁶Description of scores in ProSight Lite:

- a. **PCS**. Proteoform characterization score described in previous manuscript [6]. Range, 0–infinity. Higher numbers indicate better scores.
- b. **P-Score**. ProSight P-Score described in previous manuscript [23]. Range, 0–1. Lower numbers indicate better scores.
- c. **% Fragments Explained**. The percentage of the total number of fragments observed that are explained by the fragmentation data. This percentage can vary depending on the quality of deconvolution/deisotoping algorithm. Range: 0–100 %. Higher numbers indicate better deconvolution (less noise); lower numbers either indicate that more than one protein was isolated and fragmented simultaneously or that the deconvolution algorithm generated unrelated noise.
- d. **% Residue Cleavages**. The percentage of inter-residue bonds where at least one fragment ion was observed. Range: 0–100 %. Higher numbers indicate better scores.

for cysteine and methionine, and then click **Save**. This step prevents errors in the process of copying and pasting data from multiple sources and should be used when the protein being studied is present in the UniProt Knowledgebase.

3. Iteration 1. Results are obtained (Fig. 8) indicating poor agreement between the experimental data and the candidate sequence. The excellent P-score of 2.7×10^{-20} arising from the strong coverage of the C-terminal region of the protein indicates that the protein is identified as carbonic anhydrase II; however, the theoretical intact mass is 88.89 Da higher than the observed intact mass, and no N-terminal fragment ions are observed. At this stage, we have identified the protein, but have not fully characterized the proteoform due to the large mass error.
4. Iteration 2. To correct this mismatch between the theoretical and observed data, first select **Modify Candidate Sequence**, remove the N-terminal methionine from the sequence in the text box, and click **Save**. The results (Fig. 9) show an identical P-score (2.7×10^{-20}) and no new N-terminal fragment ions; however, the mass difference shows that the theoretical mass is now 42.15 Da lower than the observed mass.
5. Iteration 3. The final iteration requires the modification of the N-terminus with an acetyl group (with a mass of 42.01 Da). To add this modification, click on the gray “N” to the left of the sequence and, from the **Modification** box on the lower right, select “**Acetylation**.” Note that once changes are made, the coverage and scores update in real time. This feature allows for many changes to be made iteratively as the user improves the match of the theoretical proteoform to the observed data. The final result can be seen in Fig. 10. The P-score is improved to 6.5×10^{-48} , multiple fragment ions report on both termini, and the precursor mass error is reduced to just 0.14 Da or 4.9 ppm. After iteration 3, the proteoform has been correctly and fully characterized.
6. Exports can again be performed as described above in Subheading 3.2, **step 10**.

Acknowledgments

The authors are grateful to members of the Kelleher Research Group/Proteomics Center of Excellence for helpful discussions, particularly Joseph Greer, Richard LeDuc, and Bryan Early. This work was supported by Award No. P41GM108569 from the National Institute of General Medical Sciences. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

1. Erickson BK, Jedrychowski MP, McAlister GC, Everley RA, Kunz R, Gygi SP. Evaluating multiplexed quantitative phosphopeptide analysis on a hybrid quadrupole mass filter/linear ion trap/orbitrap mass spectrometer. *Anal Chem*. 2015; 87(2):1241–1249. DOI: 10.1021/ac503934f [PubMed: 25521595]
2. Beck S, Michalski A, Raether O, Lubeck M, Kaspar S, Goedecke N, Baessmann C, Hornburg D, Meier F, Paron I, Kulak NA, Cox J, Mann M. The impact II, a very high-resolution quadrupole time-of-flight instrument (QTOF) for deep shotgun proteomics. *Mol Cell Proteomics*. 2015; 14(7): 2014–2029. DOI: 10.1074/mcp.M114.047407 [PubMed: 25991688]
3. Kelleher NL. Top-down proteomics. *Anal Chem*. 2004; 76(11):197A–203A. [PubMed: 14697051]

4. LMS, NLK. Consortium for Top Down P. Proteoform: a single term describing protein complexity. *Nat Methods*. 2013; 10(3):186–187. DOI: 10.1038/nmeth.2369 [PubMed: 23443629]
5. Zheng Y, Fornelli L, Compton PD, Sharma S, Canterbury J, Mullen C, Zabrouskov V, Fellers RT, Thomas PM, Licht JD, Senko MW, Kelleher NL. Unabridged analysis of human histone H3 by differential top-down mass spectrometry reveals hypermethylated proteoforms from MMSET/NSD2 overexpression. *Mol Cell Proteomics*. 2015; doi: 10.1074/mcp.M115.053819
6. Dang X, Scotcher J, Wu S, Chu RK, Tolic N, Ntai I, Thomas PM, Fellers RT, Early BP, Zheng Y, Durbin KR, Leduc RD, Wolff JJ, Thompson CJ, Pan J, Han J, Shaw JB, Salisbury JP, Easterling M, Borchers CH, Brodbelt JS, Agar JN, Pasa-Tolic L, Kelleher NL, Young NL. The first pilot project of the consortium for top-down proteomics: a status report. *Proteomics*. 2014; 14(10):1130–1140. DOI: 10.1002/pmic.201300438 [PubMed: 24644084]
7. Tian Z, Tolic N, Zhao R, Moore RJ, Hengel SM, Robinson EW, Stenoien DL, Wu S, Smith RD, Pasa-Tolic L. Enhanced top-down characterization of histone post-translational modifications. *Genome Biol*. 2012; 13(10):R86.doi: 10.1186/gb-2012-13-10-r86 [PubMed: 23034525]
8. Barnidge DR, Dasari S, Botz CM, Murray DH, Snyder MR, Katzmann JA, Dispenzieri A, Murray DL. Using mass spectrometry to monitor monoclonal immunoglobulins in patients with a monoclonal gammopathy. *J Proteome Res*. 2014; 13(3):1419–1427. DOI: 10.1021/pr400985k [PubMed: 24467232]
9. Fornelli L, Damoc E, Thomas PM, Kelleher NL, Aizikov K, Denisov E, Makarov A, Tsybin YO. Analysis of intact monoclonal antibody IgG1 by electron transfer dissociation Orbitrap FTMS. *Mol Cell Proteomics*. 2012; 11(12):1758–1767. DOI: 10.1074/mcp.M112.019620 [PubMed: 22964222]
10. Zhang H, Cui W, Gross ML. Mass spectrometry for the biophysical characterization of therapeutic monoclonal antibodies. *FEBS Lett*. 2014; 588(2):308–317. DOI: 10.1016/j.febslet.2013.11.027 [PubMed: 24291257]
11. Ansong C, Wu S, Meng D, Liu X, Brewer HM, Deatherage Kaiser BL, Nakayasu ES, Cort JR, Pevzner P, Smith RD, Heffron F, Adkins JN, Pasa-Tolic L. Top-down proteomics reveals a unique protein S-thiolation switch in *Salmonella Typhimurium* in response to infection-like conditions. *Proc Natl Acad Sci U S A*. 2013; 110(25):10153–10158. DOI: 10.1073/pnas.1221210110 [PubMed: 23720318]
12. Cheon DH, Nam EJ, Park KH, Woo SJ, Lee HJ, Kim HC, Yang EG, Lee C, Lee JE. Comprehensive analysis of low-molecular-weight human plasma proteome using top-down mass spectrometry. *J Proteome Res*. 2015; doi: 10.1021/acs.jproteome.5b00773
13. Tran JC, Zamdborg L, Ahlf DR, Lee JE, Catherman AD, Durbin KR, Tipton JD, Vellaichamy A, Kellie JF, Li M, Wu C, Sweet SM, Early BP, Siuti N, LeDuc RD, Compton PD, Thomas PM, Kelleher NL. Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature*. 2011; 480(7376):254–258. DOI: 10.1038/nature10575 [PubMed: 22037311]
14. Liu X, Hengel S, Wu S, Tolic N, Pasa-Tolic L, Pevzner PA. Identification of ultramodified proteins using top-down tandem mass spectra. *J Proteome Res*. 2013; 12(12):5830–5838. DOI: 10.1021/pr400849y [PubMed: 24188097]
15. Zamdborg L, LeDuc RD, Glowacz KJ, Kim YB, Viswanathan V, Spaulding IT, Early BP, Bluhm EJ, Babai S, Kelleher NL. ProSight PTM 2.0: improved protein identification and characterization for top down mass spectrometry. *Nucleic Acids Res*. 2007; 35(Web Server issue):W701–W706. DOI: 10.1093/nar/gkm371 [PubMed: 17586823]
16. Cannon JR, Holden DD, Brodbelt JS. Hybridizing ultraviolet photodissociation with electron transfer dissociation for intact protein characterization. *Anal Chem*. 2014; 86(21):10970–10977. DOI: 10.1021/ac5036082 [PubMed: 25270663]
17. Fellers RT, Greer JB, Early BP, Yu X, LeDuc RD, Kelleher NL, Thomas PM. ProSight lite: graphical software to analyze top-down mass spectrometry data. *Proteomics*. 2015; 15(7):1235–1238. DOI: 10.1002/pmic.201570050 [PubMed: 25828799]
18. Bairoch A, Apweiler R, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Natale DA, O'Donovan C, Redaschi N, Yeh LS. The universal protein resource (UniProt). *Nucleic Acids Res*. 2005; 33(Database issue):D154–D159. DOI: 10.1093/nar/gki070 [PubMed: 15608167]
19. Liu X, Inbar Y, Dorrestein PC, Wynne C, Edwards N, Souda P, Whitelegge JP, Bafna V, Pevzner PA. Deconvolution and database search of complex tandem mass spectra of intact proteins: a

- combinatorial approach. *Mol Cell Proteomics: MCP*. 2010; 9(12):2772–2782. DOI: 10.1074/mcp.M110.002766 [PubMed: 20855543]
20. Carvalho PC, Xu T, Han X, Cociorva D, Barbosa VC, Yates JR 3rd. YADA: a tool for taking the most out of high-resolution spectra. *Bioinformatics*. 2009; 25(20):2734–2736. DOI: 10.1093/bioinformatics/btp489 [PubMed: 19684088]
21. Kessner D, Chambers M, Burke R, Agus D, Mallick P. ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics*. 2008; 24(21):2534–2536. DOI: 10.1093/bioinformatics/btn323 [PubMed: 18606607]
22. Roepstorff P, Fohlman J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom*. 1984; 11(11):601.doi: 10.1002/bms.1200111109 [PubMed: 6525415]
23. Meng F, Cargile BJ, Miller LM, Forbes AJ, Johnson JR, Kelleher NL. Informatics and multiplexing of intact protein identification in bacteria and the archaea. *Nat Biotechnol*. 2001; 19(10):952–957. DOI: 10.1038/nbt1001-952 [PubMed: 11581661]

Fig. 1. Example chromatogram showing the separation of five intact proteins present in TopDownStandard.raw. The MS and MS/MS spectra for ubiquitin and carbonic anhydrase were chosen near the apex of the respective peaks

The image shows a software dialog box titled "Xtract Settings". It contains several configuration options for mass spectrometry data analysis. The "Generate Masses Mode" section has two radio buttons: "M" (selected) and "MH+". The "Resolution@400" field is set to "100000.0". The "S/N Threshold" field is set to "3". The "Mass Range" section shows "Low Mass: 500.0000" and "High Mass: 2000.0000". The "Fit Factor" is set to "44" with a percentage sign, and the "Remainder" is set to "25" with a percentage sign. The "Averagine Table" dropdown menu is set to "Averagine". The "Max Charge" field is set to "40". On the right side of the dialog, there are four buttons: "Xtract", "Cancel", "Help", and "Advanced <<".

Fig. 2.

Selected Xtract parameters to be used for deconvolution and deisotoping. An explanation of these parameters may be found in Note 4

Fig. 3.
Demonstration of exporting post-Xtract data onto the clipboard for further analysis by ProSight Lite

Fig. 4. Home screen of ProSight Lite showing the two pieces of information needed to perform the analyses outlined in Subheading 3.2, **steps 2 and 6**

Experimental Data

Precursor **3.2.3**
8559.6556

Precursor Mass Type
☒ Monoisotopic ☐ Average

Mass Mode **3.2.5**
☒ M (neutral) ☐ MH+

Fragmentation Methods **Note 5**
☐ CID
☒ HCD
☐ SID
☐ ECD
☐ ETD
☐ EThcD
☐ IRMPD
☐ UVPD
☐ BY and CZ*

Fragmentation Tolerance
10 ppm

Fragments (one per line) **3.2.4**

6110.324220
6127.352323
6210.369302
6228.374662
6325.40033
6343.424054
6395.409815
6412.440663
6422.442228
6430.439334
6440.426843
6468.450974
6491.460719
6509.491103
6527.50624
6620.52017
6638.546767

Save Cancel

Fig. 5. Experimental Data panel. Precursor and fragmentation data must be provided and other parameters must be set prior to further analysis by ProSight Lite

Modify Sequence

UniProt Accession: **3.3.2**

Sequence

MQIFVKLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLLFAGKQLEDG
RTLSDYNIQKESTLHLVLRGG **3.2.7**

Fixed Modifications **3.2.8**

Cysteine

- ☒ No Modification
- ☐ S-carboxamidoethyl-L-cysteine
- ☐ S-pyridylethyl-L-cysteine
- ☐ S-carboxamidomethyl-L-cysteine
- ☐ cysteine mercaptoethanol

Methionine

- ☒ No Modification
- ☐ L-methionine sulfoxide
- ☐ L-methionine sulfone

Fig. 6. Modify Sequence panel. An amino acid sequence may be directly pasted into the sequence text box or automatically retrieved using a UniProt accession number

Fig. 7.
Successful analysis of bovine ubiquitin with ProSight Lite. Scores are discussed in detail in Note 6

Fig. 8.

Iteration 1: Partial characterization of bovine carbonic anhydrase. The protein is identified, but the proteoform (*see* Note 1 for a detailed description) is not fully characterized

Fig. 9.

Iteration 2: Improved characterization of bovine carbonic anhydrase as a result of removing the initial methionine from the sequence prior to researching

Fig. 10.

Iteration 3: Complete characterization of the N-terminal methionine cleaved, N-terminally acetylated bovine carbonic anhydrase II proteoform