
Article:
Łącki, MK, Lermyte, F, Miasojedow, B et al. (2019) masstodon: A Tool for Assigning Peaks and Modeling Electron Transfer Reactions in Top-Down Mass Spectrometry. Analytical chemistry, 91 (3). pp. 1801-1807. ISSN: 0003-2700

https://doi.org/10.1021/acs.analchem.8b01479

---

masstodon: a tool for assigning peaks and modeling electron transfer reactions in top-down mass spectrometry

Mateusz K. Łącki,[*,†] Frederik Lermyte,[*,‡,¶,§] Błażej Miasojedow,[‖] Michał P. Startek,[‖] Frank Sobott,[‡,⊥,#] Dirk Valkenborg,[¶,@,△] and Anna Gambin[‖]

†*University Medical Center, Johannes Gutenberg University, Mainz D-55131, Germany*
‡*Biomolecular and Analytical Mass Spectrometry Group, Department of Chemistry, University of Antwerp, Antwerp, Belgium*
¶*Centre for Proteomics, University of Antwerp, 2000 Antwerp, Belgium*
§*School of Engineering, University of Warwick, Coventry CV4 7AL, UK*
‖*Department of Mathematics, Informatics, and Mechanics, University of Warsaw, 02-097 Warsaw, Poland*
⊥*Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, UK*
#*School of Molecular and Cellular Biology, University of Leeds, Leeds, UK*
@*Flemish Institute for Technological Research (VITO), 2400 Mol, Belgium*
△*Interuniversity Institute for Biostatistics and Statistical Bioinformatics, Hasselt University, 3500 Hasselt, Belgium*

E-mail: matteo.lacki@uni-mainz.de; F.Lermyte@warwick.ac.uk

Abstract

Top-down mass spectrometry methods are becoming continuously more popular in the effort to describe the proteome. They rely on the fragmentation of intact protein ions inside the mass spectrometer. Among the existing fragmentation methods, electron transfer dissociation is known for its precision and wide coverage of different cleavage sites. However, several side reactions can occur under ETD conditions, including non-dissociative electron transfer and proton transfer reaction. Evaluating their extent can provide more insight into reaction kinetics as well as instrument operation. Furthermore, preferential formation of certain reaction products can reveal important structural information. To the best of our knowledge, there are currently no tools capable of tracing and analyzing the products of these reactions in a systematic way.

In this article, we present in detail `masstodon`: a computer program for assigning peaks and interpreting mass spectra. Besides being a general purpose tool, `masstodon` also offers the possibility to trace the products of reactions occurring under ETD conditions and provides insights into the parameters driving them. It is available free of charge under the GNU AGPL V3 public license.

Introduction

In recent years, there has been growing interest in electron-based dissociation (ExD) – primarily electron capture (ECD)[1] and electron transfer dissociation (ETD)[2] in protein mass spectrometry.[3] These fragmentation methods allow the cleavage of the backbone of a protein or peptide without significantly disrupting other bonds (even preserving noncovalent in-

teractions) and, as such, much effort has gone into the use of ExD methods for top-down sequencing, as well as the study of labile post-translational modifications and even binding sites of non-covalent ligands.[4–15] Additionally, considerable efforts have been made to determine preferential reaction pathways and cleavage sites in ExD of known precursors, to obtain insight into gas-phase protein/peptide conformation[16–26] as well as to investigate the reaction mechanism.[27–29] Ideally, reaction products are not only identified, but also quantified in these efforts. Because of the information-rich nature of top-down ExD spectra, data processing is usually performed with the help of specialized software.

Most of the readily available tools for the analysis of these data – e.g. `THRASH`[30], `MASH`[31,32], `DeconMSn`[33], `Decon2LS`[34] – use an averagine-scaling approach[35] to determine charge states, monoisotopic masses, and ion intensities of the molecules present in the sample. This approach allows estimation of monoisotopic mass values while only requiring limited information about the species in the spectra – specifically, the *average* composition of the type of ions present. In ETD spectra, however, one commonly observes an overlap of slightly shifted isotopic distributions, making the application of these methods somewhat less straightforward. These shifts are caused by side reactions occurring under ETD conditions, and result in ions differing in mass by one (or a few) hydrogen atoms. Without extremely high resolution, the shift by a hydrogen mass (+1.0078 Da) is indistinguishable from that by a carbon-13 isotope (+1.0034 Da), leading to the observation of a single isotope 'cluster' that is shifted and broadened compared to the expected fragment isotope distribution. These shifts and distortions, however, reveal additional information on the reaction pathways that the ions underwent inside the instrument.[26,36,37] In the (quite common) case where resolving power is sufficient to separate isotope peaks, but not to reveal the hyperfine structure, it is therefore desirable to use more exact theoretical models of the observed isotope distributions that go beyond the simple averagine approximation. This way, one can make better use of the information contained in the spectra. These reasons motivated the development of `masstodon`: a computational tool capable of tracing and quantifying ETD products in potentially high resolution MS2 spectra. Its prototype has previously been used for deconvolution of complex isotope clusters occurring in top-down ETD spectra acquired on a Waters Synapt G2 Q-IM-TOF instrument.[38] We have also shown how to infer branching ratios and that these show a high degree of correlation with collision cross-sections and gas-phase conformations of proteins, e.g. ubiquitin.[26]

Here, we present a much more advanced version of that tool and describe in detail the algorithms it makes use of. Compared to the prototype, `masstodon` can: (1) search the input mass spectra for reaction products of multiple precursor ions together with a list of arbitrary compounds, (2) accommodate the annotation of high resolution spectra using the `IsoSpec` algorithm for the generation of the theoretic isotopic envelopes[39], including the option to use either *profile* or *centroid* type data, (3) estimate the extent of different reaction pathways using also the information present in the estimated fragment intensities, and estimate the probability that an ion will undergo non-dissociative electron transfer (ETnoD), proton transfer reaction (PTR), or fragmentation (ETD with or without subsequent hydrogen abstraction). Finally, `masstodon`'s code-base has been ported to Python 3, making it maintainable, extendable, and easy to integrate on all major platforms, including Windows, Linux and MacOS. The findings are visualized using `matplotlib` and `plotly` modules. `plotly` outputs can be saved as html files and shared with other users who need only a browser to interactively explore the outcomes.

The program performs a series of steps summarized in Figure 1 and described in detail below. First, the mass spectrum is read in as one of several supported formats (see Methods). Then, a list of potential reaction products is generated based on a user-specified precursor sequence, and the theoretical isotope distribution for each of these products is calcu-

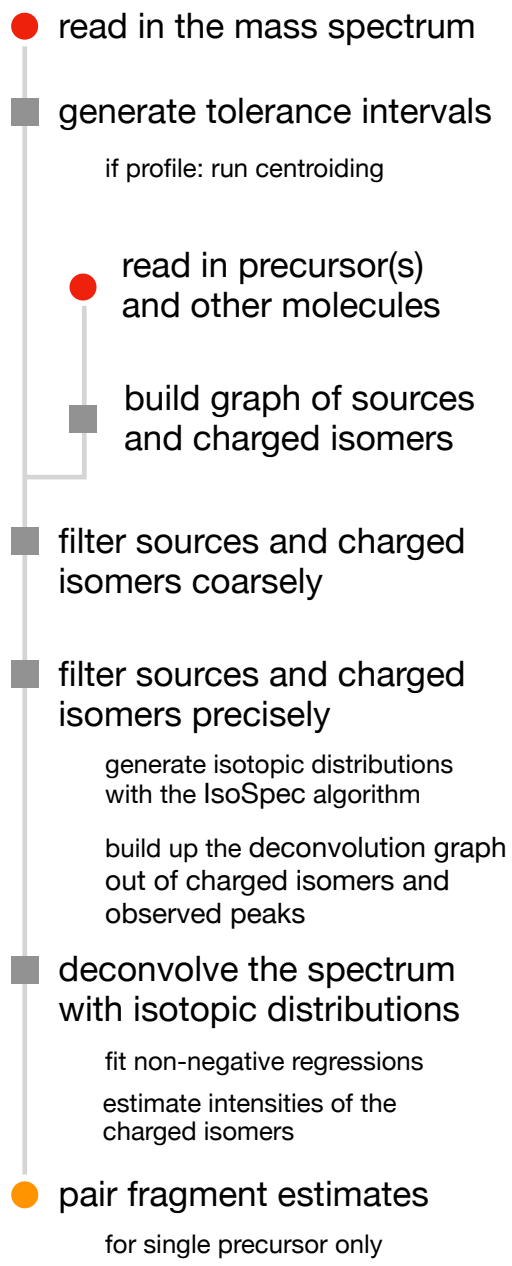

Figure 1: The workflow of the `masstodon`.

lated. The spectrum is pre-processed to make it easier to search it for theoretical isotopologues. Specifically, if profile (as opposed to centroid) data are used, this step involves the clustering of (m/z, intensity) pairs belonging to the same peak at the limits of the instrument's precision. The input molecules are arranged into a bipartite *sources-charged-isomers* graph (SI-graph for short). We call charged isomers *q-isomers* for short. The sources correspond to the user-specified precursor sequence(s), and the q-isomers to groups of MS-indistinguishable molecules that can potentially explain signals in the spectrum. During ETD, one source usually generates many potential q-isomers, among others in form of c and z ions. The next two filtering steps trim the SI-graph of all substances that do not appear in the spectrum. During coarse filtering, we discard q-isomers that have no peaks in the spectrum anywhere within the distance of three standard deviations (see Supporting Information) from their average mass to charge ratio. The precise filtering then meticulously selects q-isomers based on their precise theoretical m/z ratios obtained with the `IsoSpec` algorithm.[39]

As discussed before, signals from different q-isomers might overlap.[38] Using the theoretical isotopic distributions, we deconvolve them by means of the well-established non-negative least squares regression.[40,41] In case of a unique source of products, `masstodon` pairs the intensities of complementary fragments, enabling us to estimate intensities of particular fragmentations.

The rest of the paper is organized as follows: in *Methods* we present the stages of the proposed workflow in detail. In the *Results* section we discuss the application of the software on particular use-cases. We finish by presenting a summary and future perspective in the *Conclusions* section. The *mathematical underpinnings* together with a more precise description of the algorithms are deferred to the Supplemental Information (SI).

Methods

Spectrum. `masstodon` can read in the mass spectral data in mzXML and mzML through the use of the `Pyteomics` Python module[42]. Raw Thermo files and Waters raw folders can be exported to mzML using, for instance, MZmine2.[43] `masstodon` operates on per-scan basis. Raw, tab-separated `ASCII` files with two columns (m/z values and their corresponding intensities) are also supported.

We assume that the input spectrum is well calibrated, as recalibration within `masstodon` is currently not supported. The spectrum can be centroided or in profile mode; in the latter

case, `masstodon` executes its own centroiding procedure. To mitigate the possibility of fitting to noise peaks, we offer the user the possibility to trim signals below a user-specified threshold.

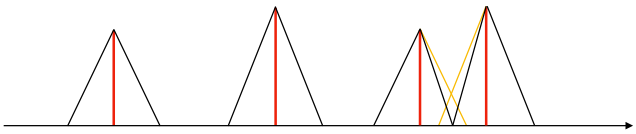

Figure 2: Tolerance intervals, as represented by the bases of triangles around the centroids (in red). We shorten the tolerance intervals to avoid their overlaps, as shown on the example of the two right-most peaks.

Even for well-calibrated data without systematic error, observed m/z ratios can be slightly off from their predicted positions due to random error, instrument drift, inheren limits to instrumental mass accuracy[44], etc. To accommodate this possibility, we set up tolerance intervals around the (centroid) peaks. The length of the interval can be precised in absolute terms, either as a fraction of a Dalton (Thompson), in milli-mass units (mmu), or in relative terms in parts per million (ppm). If two peaks are too close, the two intervals could overlap, making it harder to uniquely prescribe theoretical explanations to a given region of a spectrum. To avoid this, we shrink the overlapping intervals, so that they meet in the midway between the two observed peaks, see Figure 2.

Input molecules. The user must provide information on the potential *sources* of the signal, such as the precursor protein(s), and any other chemical molecules he wishes to search for. For each precursor protein, `masstodon` automatically generates products of the most frequent electron transfer driven reactions described in the opening section of the SI and Table 1. These reactions are assumed to occur potentially multiple times on the same molecules, leading to reaction pathways.[38] Users must specify each precursor's amino acid sequence and charge state, as this imposes a limit on how many electron or proton transfer steps the ion can possibly undergo before having its net charge completely neutralized, thereby becoming undetectable by MS. Every amino acid can be arbitrarily modified, as long as the resulting atom counts are non-negative integers. Common PTMs are included in the software.

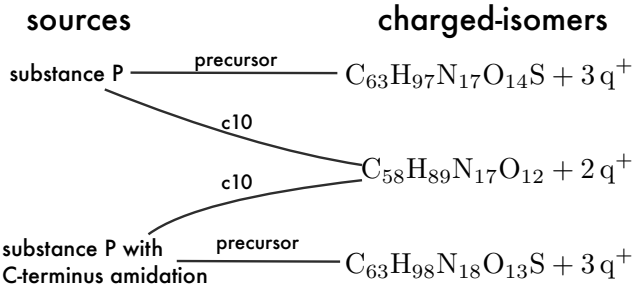

Figure 3: A small SI-graph obtained for two sources: pure triply charged Substance P and its copy with C-terminal amidation. The graph is bipartite: edges are only between sources and charged-isomers.

Molecules other than the provided precursor proteins are not subject to any reaction. Specifying those can be used to query mass spectra for the presence of an arbitrary set of molecules, each defined solely by their chemical formula and charge. One apparent use-case is to perform a search of proteins downloaded directly from the Uniprot database.[45]

All sources (precurors and *general molecules*) form one set of nodes in a bipartite graph – the SI-graph, see Figure 3. Note that it is a slight generalization of the protein-peptide graph occurring throughout proteomics.[46] The other set of nodes is formed by *charged-isomers* (q-isomers for short) that represent chemical entities that cannot be distinguished with MS. The set of q-isomers includes all reaction products of the precursor proteins and a copy of each general chemical molecule. An edge in the SI-graph denotes that a source is a particular q-isomer. Of course, multiple sources can generate the same q-isomer. Subsequent stages of the algorithm perform the trimming the SI-graph and establish the overall intensities with which q-isomers appear in the mass spectrum.

`masstodon` contains information on the masses and natural abundances of the naturally occurring isotopes of all existing elements reported by the International Union of Pure and Applied Chemistry. It is possible for the user to provide custom isotopic ratios for all the considered molecules.

The filtering. We have implemented two filters that significantly reduce the final number of q-isomers that can explain the observed mass spectrum. The first filter crudely approximates

the overall position of a q-isomers in the m/z domain by a single interval. This interval is delimited by $\langle \frac{m}{z} \rangle \pm 3\sigma(\frac{m}{z})$, where $\langle \frac{m}{z} \rangle$ is the theoretical average mass to charge ratio and $\sigma(\frac{m}{z})$ is the respective standard deviation. The SI provides a derivation of these values and the underlying rationale. `masstodon` iteratively checks if any of the above intervals intersects with any of the previously described tolerance intervals. All q-isomers that do not intersect any tolerance interval are filtered out. In other words, if no signals were detected close to the expected (average) m/z for a theoretical ion, this ion is not further considered during spectral deconvolution. The above filter is fast but imprecise.

The second layer is much more meticulous. For each of the remaining q-isomers we generate its isotopic distribution with the `IsoSpec` calculator.[39] The calculation can be accelerated by a simplifying optimization described in the SI. Ultimately, for each q-isomer we generate a series of theoretical isotopologues, described by their m/z ratio and theoretical (relative) abundances. We then check which of these isotopologues are potentially supported by the experimental data, i.e. for which isotopologues the calculated m/z value falls within one of the previously described intervals around peaks in the observed spectrum. For each q-isomer, we then calculate the proportion of the theoretical isotope distribution that has experimental support, expressed as a percentage of its total abundance. A q-isomer is retained for the rest of the analysis, if its percentage exceeds some user-defined threshold value, with a default of 80%. For a small ion (i.e. q-isomer), where one or two isotope peaks might account for virtually all of the total signal due to that ion, detection of a handful of peaks is sufficient to exceed this threshold. Conversely, for a large fragment or intact protein, for which the intensity is spread over a large number of isotope peaks, a correspondingly larger number of peaks will need to be experimentally detected above the noise level in order for this species to be retained.

Deconvolution. The overall goal of deconvolution is to approximate the shape of the observed mass spectrum by a mixture of the isotopic distributions of all q-isomers remaining after both stages of the applied filtering. The approach we take involves the construction and the analysis of a *deconvolution graph*. This is a bipartite graph, with nodes corresponding to q-isomers and tolerance intervals around an experimentally detected peaks. We connect a q-isomer to the tolerance interval with an edge if at least one of the m/z ratios in its isotopic envelope falls within that particular interval. Connected components of the deconvolution graph correspond to independent deconvolution problems, see Figure 4. Edges additionally store information on probabilities of isotopologues that should be matched with the actually observed intensities. Each deconvolution problem is solved with `Scipy`'s `nnls` implementation of the non-negative least square algorithm.[47] Similar approaches have already been argued for in the literature.[41] More details on the fitting can be found in SI.

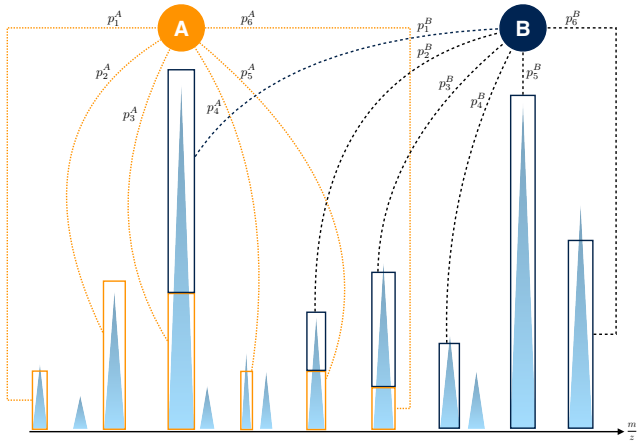

Figure 4: Relation of the deconvolution graph to an centroided mass spectrum. Circles A and B represent two different charged-isomers. The rays represent their potential presence in the explanation of the mass spectrum. The height of the triangles represents the intensities in the mass spectrum. The heights of the *fitted* rectangles are proportional to the probabilities obtained with the `IsoSpec` calculator.

Pairing of the observed ions. In general, it is complicated to infer the intensities of sources from intensities of q-isomers, the whole problem being akin to infering protein levels from measured peptide levels. For now, `masstodon` offers the possibility to study in detail the fragmentation patterns of only one source.

We have previously[26] presented a method to estimate the PTR-ETnoD branching ratios based on estimates of intensities of non-

fragmented precursor ions (see SI). We have now generalized this approach so that it can also take into account the intensities of all observed c and z fragments. To do this, we rely on pairing the intensities of fragments that can be matched. Discussion of the possible pairing strategies together with a detailed description of how they translate into optimization problems can be found in the SI. Our software can estimate the intensities of ions undergoing fragmentation via different reaction pathways and derive the probability of cleavage of each of the peptide bonds in the precursor being studied. Figure S 9 in SI presents an overview of the estimable parameters.

Results

To test the concepts behind the `masstodon`, we have analysed the previously described datasets containing spectra of purified analytes such as substance P and ubiquitin.[38,48] These datasets consist of 52 MS2 mass spectra of triply charged substance P acquired on a Waters Synapt G2 Q-IM-TOF instrument, and 4500 MS2 mass spectra of 12^+ ubiquitin acquired on a LTQ Orbitrap Velos Thermo instrument. In all cases we set an absolute precision threshold to 0.05 Th (50 mmu). Figure 5 explores `masstodon`'s outcomes resulting from the analysis of the Substance P mass spectra. Left panels show that within the typically used set of parameters `masstodon` manages to explain between 75-90% of the overall intensity (grey lines), which shows that most of the spectrum can be attributed to the considered reaction pathways. Also, the fitting quality within the regions predicted by the reaction fragments is very high, oscillating between 90 and 95% (orange lines). The middle panels plot the estimates of the distribution of potential cleavage sites for substance P. Proline fragmentation is highly unlikely and was omitted in the analysis. Our findings confirm that ETD retains similar fragmentation patterns across varying instrumental settings. The atypical worse performance for some spectra and the corresponding instability of estimates of fragmentation patterns is mainly caused by much lower abundances of fragment ions present in these data-sets. This can be seen in Figure S 13. The same effect can be noticed in the intensities attributed to individual reaction type, as shown in the right panel of Figure 5. As previously reported[48] reaction time in the ETD cell (a stacked-ring ion guide) in this instrument is minimised at high travelling-wave heights and medium wave velocities. Fewer observed ions unavoidably lead to less reliable statistics.

In order to test `masstodon` on published top-down data sets, we have analysed dataset PXD001845[49] publicly available at the ProteomeXchange, and dataset MSV000082051[50] publicly available at the MassIVE database (https://massive.ucsd.edu). Analyzed spectra differ in complexity (precursors between 18^+ up till 37^+) and the type of input mass spectrum.

The goal of project PXD001845 was to benchmark multiple fragmentation methods on an Orbitrap Fusion for top-down phosphoproteoform characterization. The project contains 183 different experiments, each with up to 20 scans, analyzing a 17.5 kDa N-terminal fragment of the mitotic regulator Bora. We analyze 16 experiments performed in ETD conditions, which amounts to analysis of 238 mass spectra. For each spectrum, we have chosen a relative precision threshold of 5ppm, which was reported by the authors of this study. Our software is currently designed to recognize only c and z fragments, so we did not analyze spectra coming from experiments where HCD fragmentation or *hybrid* methods such as EThcD or ETciD were used and in consequence b/y ions occur in a significant amount. In the analyzed experiments, the 18, 19, 22 and 24+ charge states of intact, unphosphorylated Bora and the most abundant precursor charge state (19+) of mono-, di- and tri-phosphorylated Bora were subjected to fragmentation. As shown in Figure 6, common findings are dominating, except for three cases where the `masstodon` clearly did not find as many c fragments as z. Note however, that Heck's analysis combines automatic and manual identification of fragments in spectra, whereas our analysis if fully automatic. In other words, the `masstodon`-based analysis

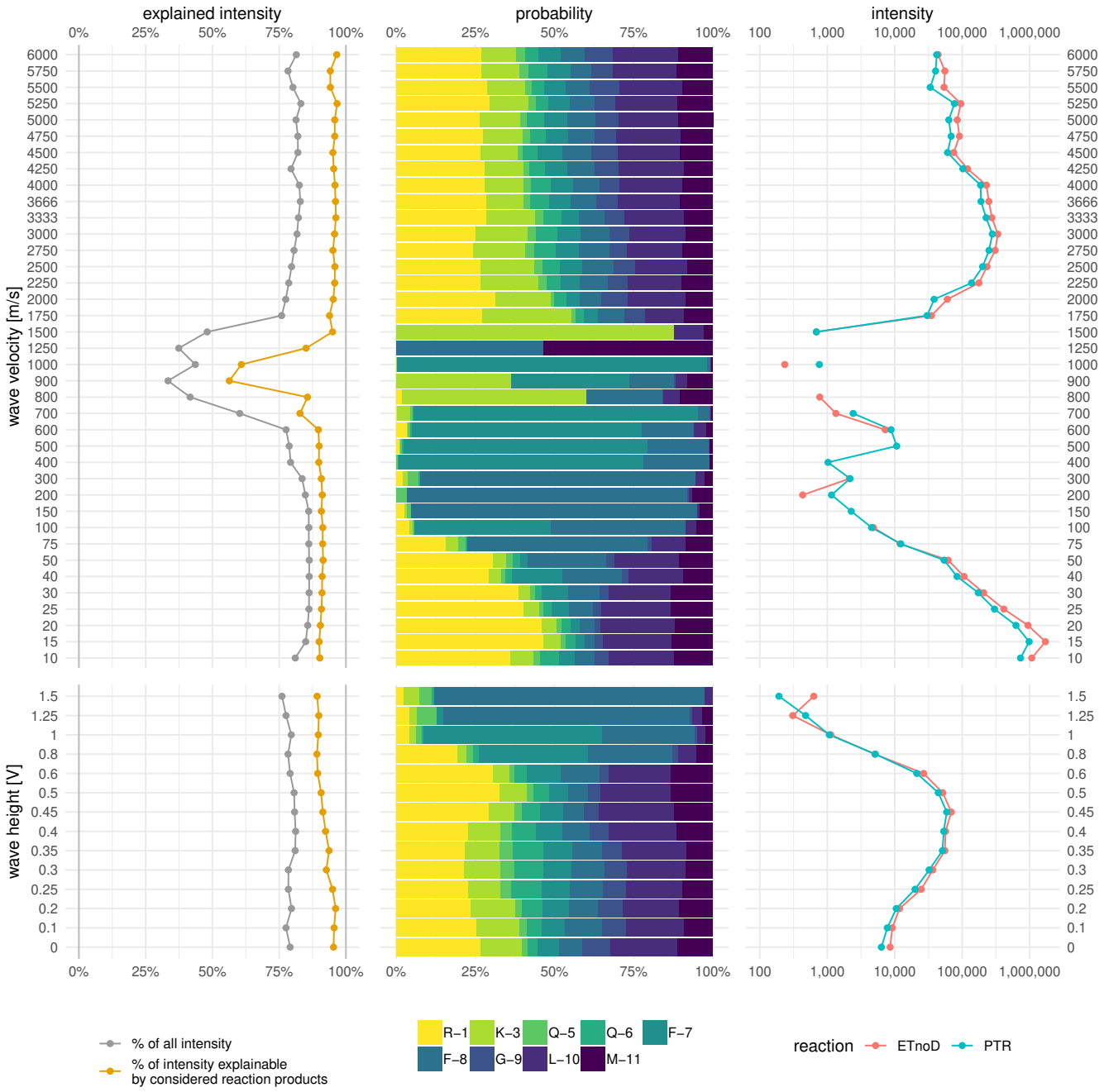

Figure 5: Results of `masstodon` fitting to mass spectra of Substance P acquired on Synapt G2. Two sets of instrumental settings were investigated: in the first we fixed wave height at 1.5 V and varied wave velocity (top panels), in the second we fixed wave velocity at 300 m/s and varied wave height (bottom panels). Left panels show how much intensity was explained by `masstodon`. The percentage of explained intensity is defined as $E(p,q) = 1 - \sum_k |p_k - q_k| / (\sum_k p_k + \sum_k q_k)$, where p and q are respectively the fitted spectrum and the experimental mass spectrum, and k iterates over different m/z values. Value of E equal one correspond to a perfect fit; on the other hand, two spectra with entirely different m/z ratios score zero. The grey line shows how much of the overall intensity in the spectrum is explained. The golden line shows how much intensity restricted only to the regions considered in the deconvolution stages can be explained, and is a quality measure we offer to assess the fitting. Middle panels show estimates of the probabilities of fragmentation along the backbone of substance P with amino acid sequence RPKPQQFFGLM. Fragmentation on prolines (P) is neglected, as it is highly unlikely due to the proline's ring structure. Right panels show estimates of the ETnoD and PTR reactions. Values in this panel are shown on a logarithmic scale.

leads to very similar results – with the added benefit that a match is required for multiple isotope peaks – in a fraction of the time of the original analysis.

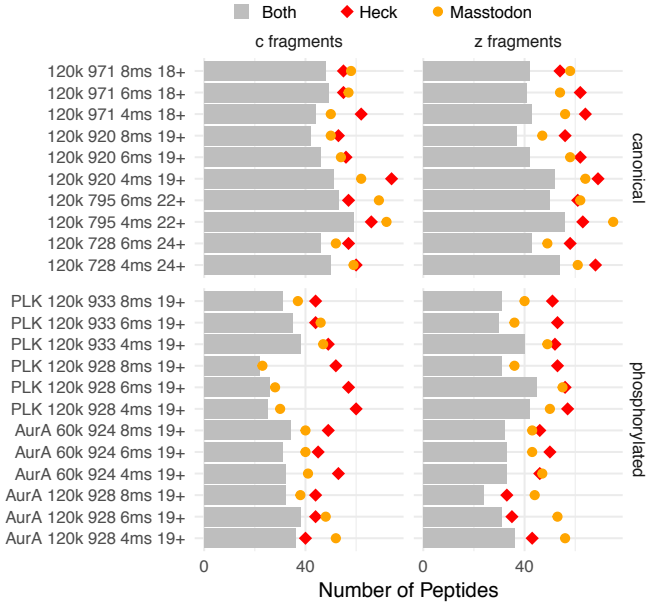

Figure 6: Comparison of findings between `masstodon` and results obtained by Heck et al.[49] Grey bars represent counts of commonly found peptides and colorful markes represent the total number of peptides found by the two methods.

The goal of project MSV000082051 was to precisely determine and characterize different proteoforms of apolipoprotein A-I (ApoA-I), all of them derived from the expression of a single gene. In as much, in the project we found additional 15 mass spectra containing signals originating from a mixture of highly charged precursor proteins. For each spectrum, we have chosen a relative precision threshold of 10ppm, which was reported by the authors of this study. Figure S 14 shows that `masstodon` can recreate most of the findings of analysts, as measured by the number of identified fragments.

Figure 7 combines all the per scan runtime of `masstodon`. It is very easy to parallize the code, as each instance of masstodon is handled by one process only. Results for the analyzed datasets were obtained in parallel, with individual runtimes varying from somewhere under a second for 3^+ small spectra, up to 10 minutes per spectrum for highly charged precursors, on a computer with 24 Intel(R) Xeon(R) 2.10GHz processors and 64GB of RAM running Linux Gentoo OS.

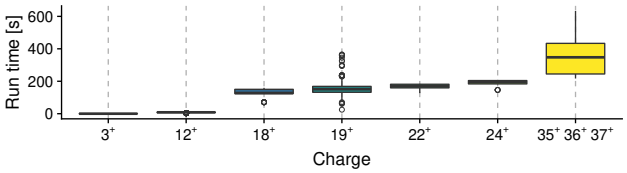

Figure 7: The run time (in seconds) of `masstodon` for different charge states. The y-axis has been square-root-transformed for clarity. Check Table S 2 for the number of spectra presented in each boxplot.

Conclusion

In this paper, we present a workflow capable of assisting analysts in processing complex high-resolution top-down protein spectra. In general, it can be used to assign high resolution mass spectra, in which the complex mixtures of isotopic structures can become an additional source of information, instead of becoming a hindrance. Moreover, `masstodon` can be used to study complex ExD mass spectra, using knowledge of gas-phase ion/ion chemistry to correctly generate the many reaction products that occur under these conditions What is more, `masstodon` offers the possibility to easily plot the findings and exchange them between users, who then do not need to go through any installation process at all. It also easily generates `CSV` files with outputs that can be directly accessed with any office computer tools. `masstodon` does not require any pre-processing (centroiding, charge deconvolution, etc.), unlike many software packages. Once the spectrum is in an acceptable format (mzXML/mzML/ASCII), our software does the rest for the user.

Finally, we would also like to stress that unlike many other programs, `masstodon` is available free of charge directly from the Python Package Index and on `github`, under https://github.com/MatteoLacki/masstodon.

Acknowledgements. We would like to thank M. A. Ciach, dr P. Dittwald, and the reviewers for their valuable comments.

Financial support. This work is supported by Polish NCN grants 2014/12/W/ST5/00592, 2015/17/N/ST6/03565, UMO-2017/26/D/ST6/00304, and partially by the Flemish SBO grant In-SPECtor, 120025, IWT. F.L. thanks the Re-

search Foundation – Flanders (FWO) for financial support.

References

(1) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. Electron capture dissociation of multiply charged protein cations. A nonergodic process. *J. Am. Chem. Soc.* **1998**, *120*, 3265–3266.

(2) Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 9528–9533.

(3) Lermyte, F.; Valkenborg, D.; Loo, J. A.; Sobott, F. Radical solutions: Principles and application of electron-based dissociation in mass spectrometry-based analysis of protein structure. *Mass Spectrom. Rev.* **2018**, *37*, 750–771.

(4) Garcia, B. A.; Shabanowitz, J.; Hunt, D. F. Characterization of histones and their post-translational modifications by mass spectrometry. *Curr. Opin. Chem. Biol.* **2007**, *11*, 66–73.

(5) Håkansson, K.; Cooper, H. J.; Emmett, M. R.; Costello, C. E.; Marshall, A. G.; Nilsson, C. L. Electron capture dissociation and infrared multiphoton dissociation MS/MS of an N-glycosylated tryptic peptide to yield complementary sequence information. *Anal. Chem.* **2001**, *73*, 4530–4536.

(6) Ayaz-Guner, S.; Zhang, J.; Li, L.; Walker, J. W.; Ge, Y. In vivo phosphorylation site mapping in mouse cardiac troponin I by high resolution top-down electron capture dissociation mass spectrometry: Ser22/23 are the only sites basally phosphorylated. *Biochemistry* **2009**, *48*, 8161–8170.

(7) Ge, Y.; Rybakova, I. N.; Xu, Q.; Moss, R. L. Top-down high-resolution mass spectrometry of cardiac myosin binding protein C revealed that truncation alters protein phosphorylation state. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 12658–12663.

(8) Tsybin, Y. O.; Fornelli, L.; Stoermer, C.; Luebeck, M.; Parra, J.; Nallet, S.; Wurm, F. M.; Hartmer, R. Structural analysis of intact monoclonal antibodies by electron transfer dissociation mass spectrometry. *Anal. Chem.* **2011**, *83*, 8919–8927.

(9) Fornelli, L.; Damoc, E.; Thomas, P. M.; Kelleher, N. L.; Aizikov, K.; Denisov, E.; Makarov, A.; Tsybin, Y. O. Analysis of intact monoclonal antibody IgG1 by electron transfer dissociation Orbitrap FTMS. *Mol. Cell. Proteomics* **2012**, *11*, 1758–1767.

(10) Cournoyer, J. J.; Pittman, J. L.; Ivleva, V. B.; Fallows, E.; Waskell, L.; Costello, C. E.; O'Connor, P. B. Deamidation: Differentiation of aspartyl from isoaspartyl products in peptides by electron capture dissociation. *Protein Sci.* **2005**, *14*, 452–463.

(11) Li, X.; Lin, C.; O'Connor, P. B. Glutamine deamidation: differentiation of glutamic acid and γ-Glutamic Acid in peptides by electron capture dissociation. *Anal. Chem.* **2010**, *82*, 3606–3615.

(12) Xie, Y.; Zhang, J.; Yin, S.; Loo, J. A. Top-down ESI-ECD-FT-ICR mass spectrometry localizes noncovalent protein-ligand binding sites. *J. Am. Soc. Mass Spectrom.* **2006**, *128*, 14432–14433.

(13) Jackson, S. N.; Dutta, S.; Woods, A. S. The use of ECD/ETD to identify the site of electrostatic interaction in noncovalent complexes. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 176–179.

(14) Yin, S.; Loo, J. A. Elucidating the site of protein-ATP binding by top-down mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 899–907.

(15) Göth, M.; Lermyte, F.; Schmitt, X. J.; Warnke, S.; von Helden, G.; Sobott, F.; Pagel, K. Gas-phase microsolvation of ubiquitin: investigation of crown ether complexation sites using ion mobility-mass spectrometry. *Analyst* **2016**, *141*, 5502–5510.

(16) Breuker, K.; Oh, H.; Horn, D. M.; Cerda, B. A.; McLafferty, F. W. Detailed unfolding and folding of gaseous ubiquitin ions characterized by electron capture dissociation. *J. Am. Chem. Soc.* **2002**, *124*, 6407–6420.

(17) Oh, H.; Breuker, K.; Sze, S. K.; Ge, Y.; Carpenter, B. K.; McLafferty, F. W. Secondary and tertiary structures of gaseous protein ions characterized by electron capture dissociation mass spectrometry and photofragment spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 15863–15868.

(18) Skinner, O. S.; McLafferty, F. W.; Breuker, K. How ubiquitin unfolds after transfer into the gas phase. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 1011–1014.

(19) Skinner, O. S.; Breuker, K.; McLafferty, F. W. Charge site mass spectra: conformation-sensitive components of the electron capture dissociation spectrum of a protein. *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 807–810.

(20) Zhang, H.; Cui, W.; Wen, J.; Blankenship, R. E.; Gross, M. L. Native electrospray and electron-capture dissociation FTICR mass spectrometry for top-down studies of protein assemblies. *Anal. Chem.* **2011**, *83*, 5598–5606.

(21) Zhang, H.; Cui, W.; Gross, M. L. Native electrospray ionization and electron-capture dissociation for comparison of protein structure in solution and the gas phase. *Int. J. Mass Spectrom.* **2013**, *354*, 288–291.

(22) Zhang, Z.; Browne, S. J.; Vachet, R. W. Exploring salt bridge structures of gas-phase protein ions using multiple stages of electron transfer and collision induced dissociation. *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 604–613.

(23) Lermyte, F.; Konijnenberg, A.; Williams, J. P.; Brown, J. M.; Valkenborg, D.; Sobott, F. ETD allows for native surface mapping of a 150 kDa noncovalent complex on a commercial Q-TWIMS-TOF instrument. *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 343–350.

(24) Lermyte, F.; Sobott, F. Electron transfer dissociation provides higher-order structural information of native and partially unfolded protein complexes. *Proteomics* **2015**, *15*, 2813–2822.

(25) Zhang, Y.; Cui, W.; Wecksler, A. T.; Zhang, H.; Molina, P.; Deperalta, G.; Gross, M. L. Native MS and ECD characterization of a fab–antigen complex may facilitate crystallization for X-ray diffraction. *J. Am. Soc. Mass Spectrom.* **2016**, *27*, 1139–1142.

(26) Lermyte, F.; Łącki, M. K.; Valkenborg, D.; Gambin, A.; Sobott, F. Conformational space and stability of ETD charge reduction products of ubiquitin. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 69–76.

(27) Tureček, F. N Cα Bond Dissociation Energies and Kinetics in Amide and Peptide Radicals. Is the Dissociation a Nonergodic Process? *J. Am. Chem. Soc.* **2003**, *125*, 5954–5963.

(28) Tureček, F.; Syrstad, E. A. Mechanism and energetics of intramolecular hydrogen transfer in amide and peptide radicals and cation-radicals. *J. Am. Chem. Soc.* **2003**, *125*, 3353–3369.

(29) Chung, T. W.; Tureček, F. Backbone and side-chain specific dissociations of z ions from non-tryptic peptides. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 1279–1295.

(30) Horn, D. M.; Zubarev, R. A.; McLafferty, F. W. Automated reduction and interpretation of high resolution electrospray mass spectra of large molecules. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 320–332.

(31) Guner, H.; Close, P. L.; Cai, W.; Zhang, H.; Peng, Y.; Gregorich, Z. R.; Ge, Y. MASH Suite: a user-friendly and versatile software interface for high-resolution mass spectrometry data interpretation and visualization. *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 464–470.

(32) Cai, W.; Guner, H.; Gregorich, Z. R.; Chen, A. J.; Ayaz-Guner, S.; Peng, Y.; Valeja, S. G.; Liu, X.; Ge, Y. MASH Suite Pro: A comprehensive software tool for top-down proteomics. *Mol. Cell. Proteomics* **2016**, *15*, 703–714.

(33) Mayampurath, A. M.; Jaitly, N.; Purvine, S. O.; Monroe, M. E.; Auberry, K. J.; Adkins, J. N.; Smith, R. D. DeconMSn: a software tool for accurate parent ion monoisotopic mass determination for tandem mass spectra. *Bioinformatics* **2008**, *24*, 1021–1023.

(34) Jaitly, N.; Mayampurath, A.; Littlefield, K.; Adkins, J. N.; Anderson, G. A.; Smith, R. D. Decon2LS: An open-source software package for automated processing and visualization of high resolution mass spectrometry data. *BMC Bioinf.* **2009**, *10*, 87.
(35) Senko, M. W.; Beu, S. C.; McLafferty, F. W. Determination of monoisotopic masses and ion populations for large biomolecules from resolved isotopic distributions. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 229–233.
(36) O'Connor, P. B.; Lin, C.; Cournoyer, J. J.; Pittman, J. L.; Belyayev, M.; Budnik, B. A. Long-lived electron capture dissociation product ions experience radical migration via hydrogen abstraction. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 576–585.
(37) Tsybin, Y. O.; He, H.; Emmett, M. R.; Hendrickson, C. L.; Marshall, A. G. Ion activation in electron capture dissociation to distinguish between N-terminal and C-terminal product ions. *Anal. Chem.* **2007**, *79*, 7596–7602.
(38) Lermyte, F.; Łącki, M. K.; Valkenborg, D.; Baggerman, G.; Gambin, A.; Sobott, F. Understanding reaction pathways in top-down ETD by dissecting isotope distributions: A mammoth task. *Int. J. Mass Spectrom.* **2015**, *390*, 146–154.
(39) Łącki, M. K.; Startek, M.; Valkenborg, D.; Gambin, A. IsoSpec: Hyperfast Fine Structure Calculator. *Anal. Chem.* **2017**, *89*, 3272–3277.
(40) Lawson, C. L.; Hanson, R. J. *Solving least squares problems*; Siam, 1995; Vol. 15.
(41) Slawski, M.; Hussong, R.; Tholey, A.; Jakoby, T.; Gregorius, B.; Hildebrandt, A.; Hein, M. Isotope pattern deconvolution for peptide mass spectrometry by non-negative least squares/least absolute deviation template matching. *BMC Bioinf.* **2012**, *13*, 291.
(42) Goloborodko, A. A.; Levitsky, L. I.; Ivanov, M. V.; Gorshkov, M. V. Pyteomics—a Python framework for exploratory data analysis and rapid software prototyping in proteomics. *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 301–304.
(43) Pluskal, T.; Castillo, S.; Villar-Briones, A.; Orešič, M. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinf.* **2010**, *11*, 395.
(44) Dittwald, P.; Valkenborg, D.; Claesen, J.; Rockwood, A. L.; Gambin, A. On the Fine Isotopic Distribution and Limits to Resolution in Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2015**, *26*, 1732–1745.
(45) Bateman, A. et al. UniProt: a hub for protein information. *Nucleic Acids Res.* **2015**, *43*, 204–212.
(46) Serang, O.; Käll, L. Solution to statistical challenges in proteomics is more statistics, not less. *J. Proteome Res.* **2015**, *14*, 4099–4103.
(47) Jones, E.; Oliphant, T.; Peterson, P. SciPy: Open source scientific tools for Python. 2001–.
(48) Lermyte, F.; Verschueren, T.; Brown, J. M.; Williams, J. P.; Valkenborg, D.; Sobott, F. Characterization of top-down ETD in a travelling-wave ion guide. *Methods* **2015**, *89*, 22–29.
(49) Brunner, A. M.; Lössl, P.; Liu, F.; Huguet, R.; Mullen, C.; Yamashita, M.; Zabrouskov, V.; Makarov, A.; Altelaar, A. M.; Heck, A. J. Benchmarking multiple fragmentation methods on an orbitrap fusion for top-down phosphoproteoform characterization. *Anal. Chem.* **2015**, *87*, 4152–4158.
(50) others„ et al. A Targeted, Differential Top-Down Proteomic Methodology for Comparison of ApoA-I Proteoforms in Individuals with High and Low HDL Efflux Capacity. *J. Proteome Res.* **2018**, *17*, 2156–2164.