There be Dragons: the dangers associated with assigning internal fragment ions of proteins

David PA Kilgour[1], Harry Taylor[1], Yury Tsybin[2], David Clarke[3], Logan Mackay[3], Luca Fornelli[4]

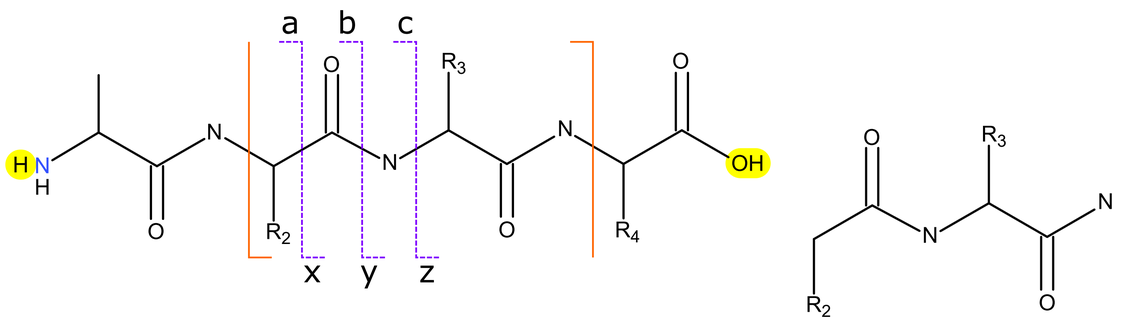
1 Nottingham Trent University; 2 Spectroswiss Sàrl; 3 University of Edinburgh; 4 University of Oklahoma

Introduction

There is a great deal of current interest in the potential utility of internal fragments generated during top-down tandem mass spectrometry of intact proteins. But, there are a wide variety of risks associated with attempting to make use of these internal fragment assignments for samples that are either potentially impure or are not already very well characterised, and, as there are so many more possible internal fragments, the value in the assignments appears to be orders of magnitude lower than for terminal fragments. Finally, there are differences between the various nomenclatures used to describe internal fragment ions, and this lack of consistency makes it harder for us, as a community, to discuss these issues as this potentially exciting area develops.

Internal fragments

Conventionally, terminal fragments are used in top-down protein analyses. Internal fragments are those created by breaking the peptide backbone in at least two places, to produce a fragment from the interior of the sequence.



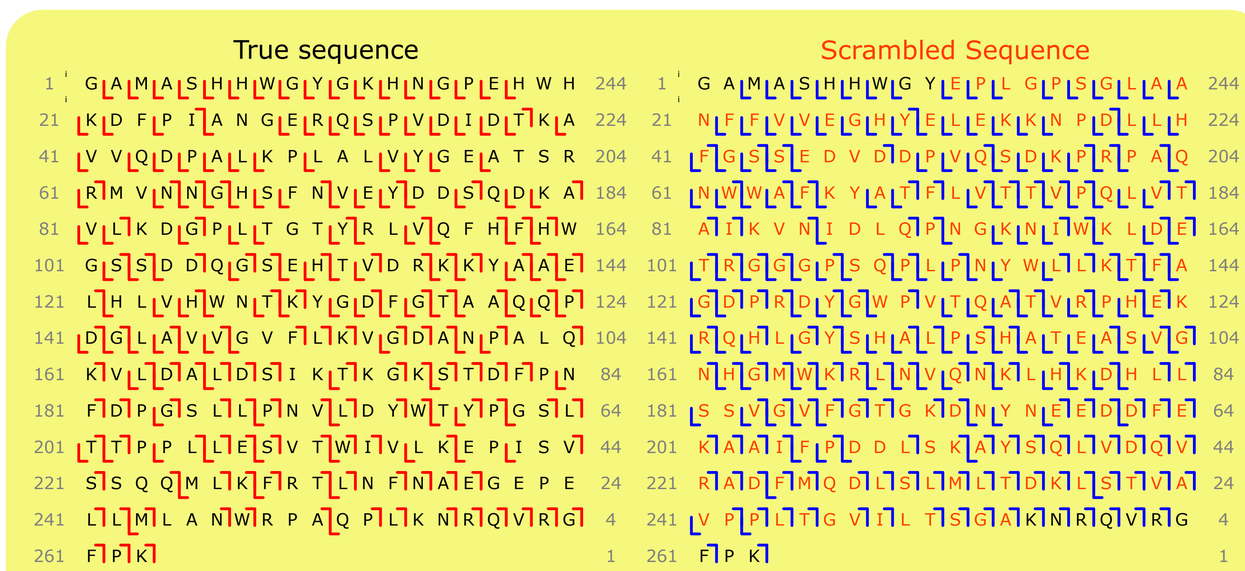
A numbers game

The main cause of the difficulties involved in making use of internal fragments is down to huge numbers of potential fragments. Take Bovine Carbonic Anhydrase as an example. For this 263 residue protein, there are 525 possible terminal fragments, but over 33k possible internal fragments.

As internal fragments are presumed to be caused by a terminal fragment being fragmented a second time, we assume that the ends of the internal fragment can mirror the possibilites for the terminal fragments in the same data. The total variety of internal fragments is the product of the numbers of N and C terminal fragment types. Therefore, if your fragmentation method produces only one major type of N and C terminal framents, then you have only one type of internal fragment to find. But, if you have 3x N terminal fragment types and 3x C terminal fragment types then you have 9 times as many possible internal fragment classes to search - which may each appear in many different charge states.

The libraries become so dense that spurious hits are almost inevitable. It may require mass accuracy levels that are beyond the routine performance levels of current technology to permit their routine use.

This can be easily demonstrated by assigning the same spectrum against internal fragment libraries that match the true sequence and against libraries built from false sequences. Commonly we see very similar internal fragment assignment statistics against false or scrambled sequences - which greatly reduces confidence in assignments. Even for FT-ICR MS data.



Showing internal fragments assigned for ECD fragments of Bovine carbonic anhydrase (22+ charge state) against the true sequence (left) and a randomly scrambled sequence (right) - where the 10 terminal residues are left in their normal sequence.

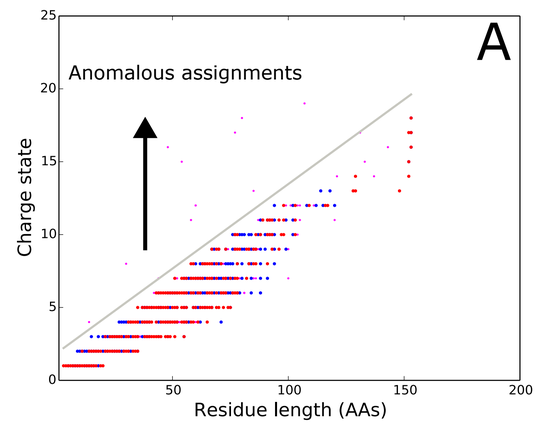
Improving confidence

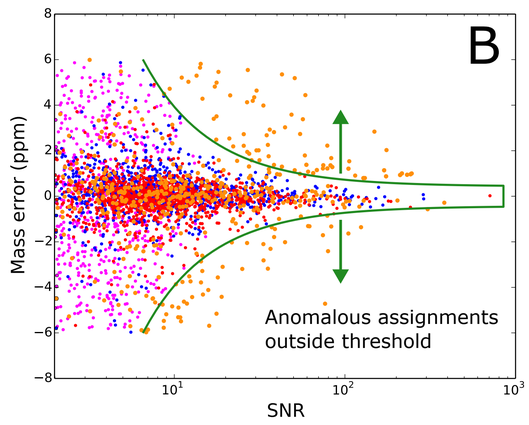
There are several approaches that can be used to improve confidence in the assigment of fragments, including internal fragments.

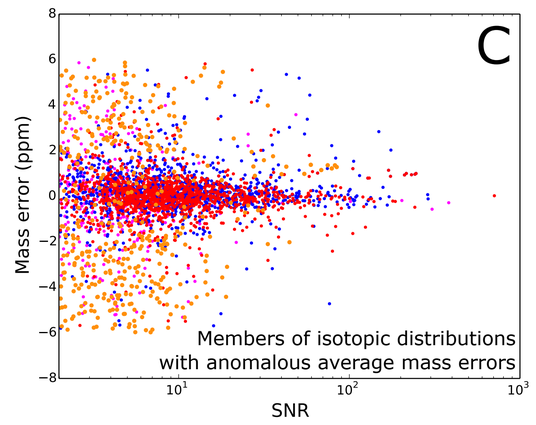
- Basic methods
  - no charge state deconvolution - mass accuracy is paramount
  - ppm error cut-off - tighter the better & 5ppm is not good enough
  - isotopic fit
- More advanced filters
  - charges per amino acid (length of fragment)
  - statistical tests - S/N vs mass error ("Prosaic") and average isotopic mass error ("Cookie Cutter")

All of these methods have been implmented in the AutoVectis Suite top-down assignment tool "AutoSeequer".

Demonstrating filtering of assignments, including internal fragments, based on looking for anomalous combinations of fragment charge state versus length [A]; S/N versus mass error [B]; and average mass error across an isotopic distribution [C] provides useful increased confidence. But, while these are very successful for terminal fragments, the confidence of internal fragments remains problematic.



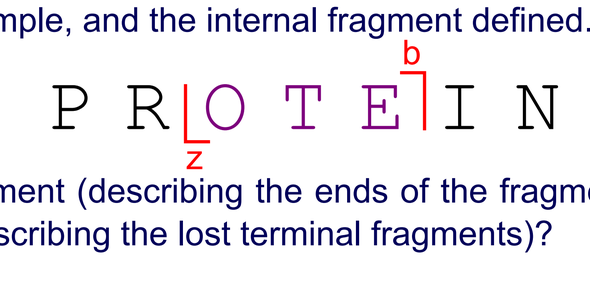




Nomenclature

There is also an issue with the lack of standardization for the nomenclature used to define internal fragments of proteins. All tools we have reviewed agree on the numbering of internal fragments. However, there is less consistency when it comes to the classification - which affects the ionic formulae. Some software, e.g.ClipsMS[(1)], defines the fragments according to what has been lost from each end of the internal fragment whereas other software, including AutoVectis, defines the fragments by what is present in the internal fragment. This adds an additional level of complexity when comparing results produced by different processing pipelines and is something that the community may wish to address.

Consider this example, and the internal fragment defined.



Is this a b&z fragment (describing the ends of the fragment detected or c&y type fragment (describing the lost terminal fragments)?

Conclusions

Internal fragments offer some potential to assist in top-down protein characterization - but there are many problems to overcome before they may become robustly applicable.

Acknowledgments

This research project has received funding from the European Horizon 2020 research and innovation program under grant agreement No 829157.

References

1. Lantz C et al. ClipsMS: An Algorithm for Analyzing Internal Fragments JPR. 2021 Mar 2;20(4):1928-35.