In-Source Fragmentation and the Sources of Partially Tryptic Peptides in Shotgun Proteomics

Overview
- This study investigates the impact of in-source fragmentation on peptide identifications from a standard protein mixture and two complex biological samples.
- In-source fragments are distinguished from endogenous tryptic peptides based on elution time criteria.

Introduction
- In shotgun proteomics, numerous partially tryptic peptides are often observed, though ideally peptides are supposed to be fully tryptic.
- The partially tryptic peptides from chymotrypsin contamination and trypsin autohydrolysis issues were largely alleviated by the use of sequencing grade modified trypsin.
- The other major sources for generating partially tryptic peptides are endogenous proteolytic activity and fragmentation whereby tryptic peptides are fragmented within the electrospray ionization source.
- To clarify the sources of partially tryptic peptides and the specificity of trypsin, we investigated the impact of in-source fragmentation on peptide identifications from a standard protein mixture and two complex biological samples (mouse brain and mouse plasma).
- In-source fragments were distinguished from endogenous tryptic peptides based on elution time criteria.

Methods
- Three different peptide samples were analyzed using LC-MS/MS (Dionex mass spectrometer): 1) a mixture of six protein standards (bovine calreticulin, bovine beta-lactoglobulin, Z. m. coli-bacteptacterium; equine skeletal muscle; bovine serum albumin; and cytochrome c); 2) mouse brain; and 3) mouse plasma.
- Parameters for ion optics were automatically tuned by Xcalibur Tune Plus to maximize ion gain. Tuning parameters were: ESI 2.2 kV, heated capillary temperature 200 °C, capillary voltage 30 V, and tube lens voltage 80 V.
- A false discovery rate of 0.05% for peptide identification was achieved by filtering peptides, using an MS-GF+ score threshold in combination with a 10 ppm mass accuracy cutoff.
- The first observed scan number for all peptides identified was plotted against predicted normalized elution times (NET) using in-house NET prediction tool (publicly available at http://proteogenomics.pnl.gov).

Results
- Partially tryptic peptides from in-source fragmentation

Conclusions
- Our results indicate in-source fragmentation can be a major source of partially tryptic peptides in LC-MS/MS proteomics and that the impact of these artifacts is much less significant for data-dependent measurements of more complex mixtures due to the larger extent of under-sampling of low-abundance species.
- It appears that the impact of trypsin-induced partially tryptic peptides is minimal (< 0.5%).
- The majority of the in-source fragments showed much later elution times compared to the predicted peptides which is consistent with a larger size of their actual parent ions. By excluding such in-source fragmentation artifacts, one can then much more effectively identify in-vivo or in-vitro proteolytic products, as well as better characterize protease activities.

Acknowledgements

References

CONTACT: Jong-Seo Kim, Ph.D. Institute for Basic Science Seoul National University, South Korea E-mail: jongseokim@ibs.re.kr

Weijun Qian, Ph.D. Biological Sciences Division Pacific Northwest National Laboratory E-mail: Weijun.Qian@pnnl.gov

Fig. 3. Relative portions of each type of tryptic peptide at the unique peptide level in the three biological samples.

Fig. 4. Sequence motifs around cleavage sites in isoelectric focusing fragments from mouse brain (A) and mouse plasma (B). Sequence motifs of partially tryptic peptides after excluding in-source artifacts from mouse brain (C) and mouse plasma (D).

Table of in-source fragmentation-generated peptides

| Type of peptide | Description | NL: 1.48E5 | NL: 1.50E5 | NL: 1.51E5 | NL: 1.65E5 | NL: 2.72E5 | NL: 3.17E4 | NL: 4.74E4 | NL: 7.27E5 | NL: 9.28E4 | NL: 1.46E5 | NL: 4.46E4 | NL: 4.92E6 | NL: 5.08E6 | NL: 5.08E6 | NL: 5.08E6 | NL: 8.81E7 | NL: 8.81E7 | NL: 8.81E7 | NL: 8.81E7 |
|-----------------|-------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| B-ion type      |             |           |           |           |           |           |           |           |           |           |           |           |           |           |           |           |           |           |           |           |           |
| Y-ion type      |             |           |           |           |           |           |           |           |           |           |           |           |           |           |           |           |           |           |           |           |           |           |
| [K/R].P type    |             |           |           |           |           |           |           |           |           |           |           |           |           |           |           |           |           |           |           |           |           |           |
| In-source fragment |             |           |           |           |           |           |           |           |           |           |           |           |           |           |           |           |           |           |           |           |           |           |
| Linear (Fully tryptic) |             |           |           |           |           |           |           |           |           |           |           |           |           |           |           |           |           |           |           |           |           |           |

Fig. 1. Extracted ion chromatograms of AVQDPAI/AIPLAV/GAIS/ATSR in a standard protein mixture (A) and in mouse brain (B) and mouse plasma (C) in mouse brain tryptic and with their in-source fragmentations in source fragments. 1A. Signal intensity of the most intense peak in view.

Fig. 2. Plots of first scan number of identified peptides vs. predicted normalized elution time (NET) in isoelectric focusing (A, mouse brain (B), and mouse plasma (C)). The observed elution times of in-source fragments appear later than predicted peaks in source fragments have the same elution times as their parent peaks.