thermoscientific

Developments in Real-Time Search on an Orbitrap Tribrid Mass Spectrometer

Jesse D. Canterbury, William D. Barshop, Graeme C. McAlister, Tony Zhao, Aaron M. Robitaille, and Romain Huguet, Thermo Fisher Scientific, San Jose, CA, 95134

ABSTRACT

Purpose: We extend the capabilities of Real-Time Search (RTS) to include: (1) filtering search results based on a multi-dimensional score using a real-time assessment of false discovery rate as a metric for setting the score threshold; (2) a "close-out" option which allows rejection of PSMs that match a single protein above a user-defined number of occurrences; (3) support for precursor and fragment neutral loss.

Methods: Instrument control software was modified with the above features, and tested with TMT-labeled triple-knockout yeast samples.

Results: Real-time search with multi-feature filtering, close-out, and neutral loss support can enhance the sensitivity and speed of proteomics analyses.

INTRODUCTION

Recent developments in instrument control architecture and processor speed on ordinary desktop computers have resulted in systems fast enough for real-time processing of mass spectrometry data. The most impactful application of these advances has been the ability to do spectral database searching in real time. Obtaining search results online offers the ability to make decisions for subsequent analysis on-line during the chromatographic elution, including which MS2 fragments to interrogate for further fragmentation [1]. In this work, we extended the capabilities of real-time search (RTS) on Thermo Scientific™ Orbitrap Eclipse™ mass spectrometers to include filtering PSMs based on real-time computation of false-discovery rate, the option to "close out" previously matched proteins, and support for neutral loss analysis.

MATERIALS AND METHODS

Sample and Method

LC-MS. Using a Thermo Scientific™ Easy-nLC™ 1200, we ran LC gradients of about 2 hours in length, ramping buffer B (95:4.9:0.1 acetonitrile:water:formic acid) from 5% to 35%. 500 ng of peptides originating from the Thermo Scientific™ Pierce™ TMT11plex Yeast Digest Standard were loaded onto Thermo Scientific™ EASY-Spray™ columns coupled via a nanoLC source to an Orbitrap Eclipse Tribrid mass spectrometer.

Real-Time Search (RTS). For real-time interpretation of MS/MS spectra, we used a version of the Comet search engine [2], derived from Sequest and optimized for performance. Spectra are packaged by the mass spectrometer's embedded software, and sent to the host PC for processing by Comet. The results are then packaged and returned to the embedded software, where results are filtered according to score criteria, and matching peaks are selected for subsequent synchronous precursor selection (SPS) and MS3 fragmentation.

Data Analysis

For offline analysis, RAW files were analyzed with Thermo Scientific™ Proteome Discoverer™ (PD) 2.4 software. Further custom analyses were carried out using code written in Lua or C#.

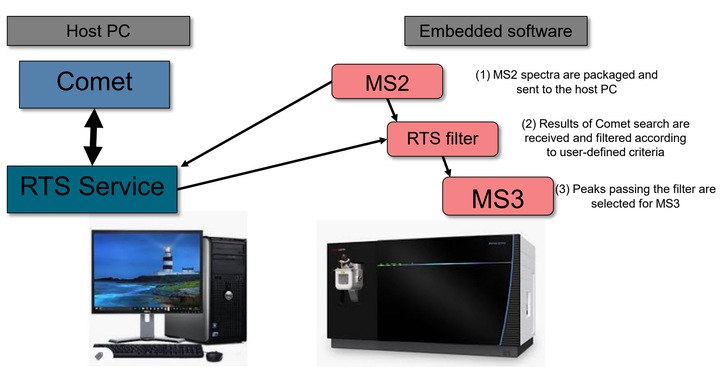
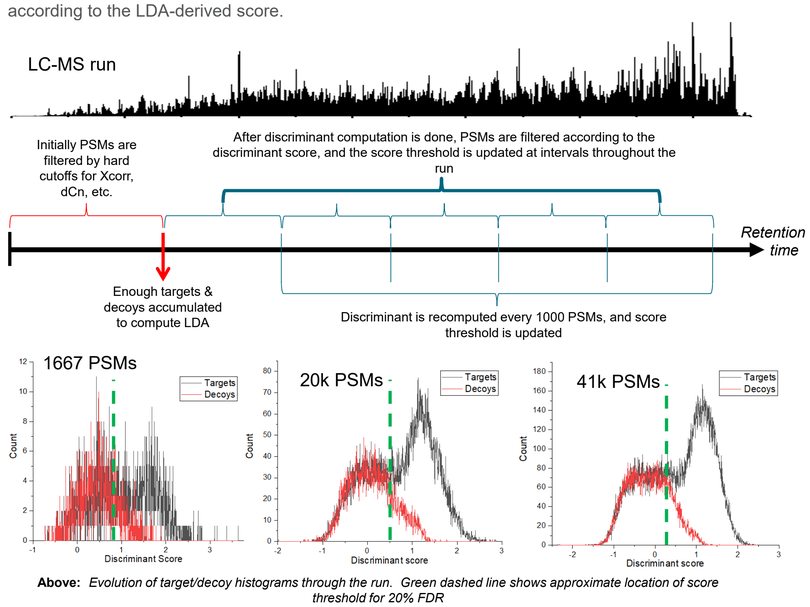


Figure 1. Overview of Real-Time Search (RTS) on an Orbitrap Eclipse Tribrid mass spectrometer.

RESULTS

Intra-run Filtering of PSMs based on FDR estimation in real time

Following an approach described by Schweppe et al. [1], we reduce a set of scores and metrics to a single score, using linear discriminant analysis (LDA). For computing the LDA we used the Accord.NET Framework, an extremely versatile library for machine learning [3]. Our feature set contains results of the Comet search as well as other separate features: Xcorr, dCn, precursor mass accuracy, fraction of ions matched, fraction of ion current explained, peptide length, and charge state. During an LC-MS run we accumulate target and decoy PSMs up to a specified minimum, compute the LDA, score all PSMs according to feature weights determined by the LDA, and then determine the score threshold needed to maintain a 20% false discovery rate. This process is repeated every 1000 target PSMs, and the score threshold is updated. Between updates, new PSMs are no longer filtered according to hard cutoffs of Xcorr, dCn, etc., but instead are filtered according to the LDA-derived score.



Above: *Evolution of target/decoy histograms through the run. Green dashed line shows approximate location of score threshold for 20% FDR*

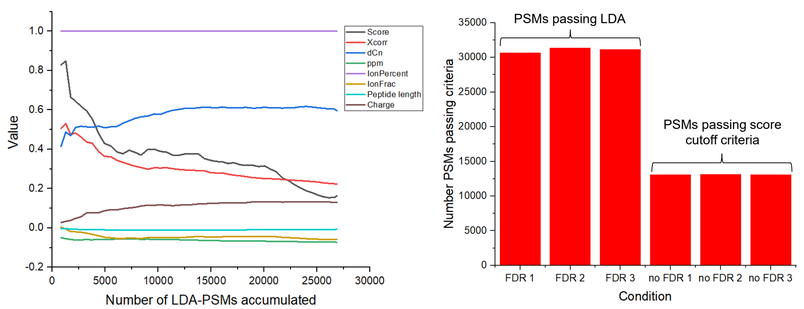
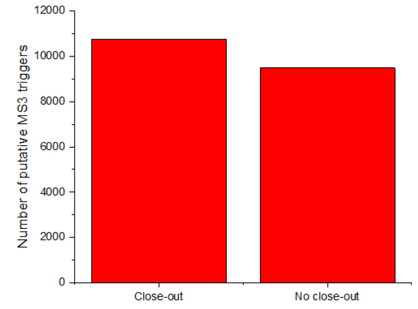


Figure 2. Description and working example of the algorithm. Target and decoy PSMs are accumulated throughout the run. Target PSMs passing a filter of Xcorr > 1 and precursor mass accuracy < 5 ppm are used as the positive inputs for discriminant calculation; targets and decoys not passing these filters are used as the negative inputs. The score threshold is calculated from the target/decoy histograms, at a target false discovery rate (FDR) of 20%. The FDR is calculated as the number of decoys divided by the number of targets. A high FDR is chosen to increase the likelihood of additional PSMs passing the filter. Whether passing PSMs translate to increased MS3 and quantitative yield depends on downstream filtering of MS2 fragment spectra.

Protein Close-out

To allow analyses to dig deeper into the proteome, we implemented the option to maintain a list of previously identified proteins. Further spectra matching peptides belonging to a protein on the list, subject to a user-defined number of maximum allowed PSMs-per-protein, are rejected by the RTS filter. The current implementation is for intra-run close-out only.

Figure 3. Triplicate LC-MS runs show an increase of about 10% in predicted MS3 triggers when using close-out vs. not using close-out. The utility of the method depends on the sample and LC conditions, and is expected to be more useful for multi-fractionated samples where eliminating matches to common cross-fraction proteins will have more of an effect on the final outcome.

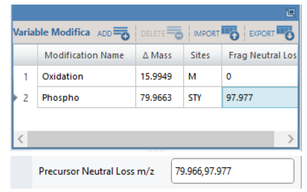


Neutral loss support

While data are not shown on this poster, further enhancements to Real-time Search enable support for precursor and fragment neutral loss capabilities that are included in the Comet search engine.

Figure 4. A screen shot of the method editor controls is shown below. For more information, please see the poster referenced below.

Please see poster TP457:
"Leveraging the extended instrument capabilities of a Tribrid MS using real-time PTM localization"
William Barshop *et al.*



CONCLUSIONS

- Real-time Search (RTS) with FDR filtering and close-out can be an effective filter for high-confidence spectrum matches, leading to more efficient use of mass spectrometer time.
- Search times in the low milliseconds and multithreaded programming ensure that real-time search doesn't interfere with other operations
- Neutral-loss support further enhances the capabilities of peptide identification in real time.

REFERENCES

1. Schweppe et al., *J. Proteome Res.* 2020, **19**, 2026-2034; Erickson et al., *J. Proteome Res.* 2019, **18**, 1299-1306.
2. Eng et al., *JASMS* 2015, **26**, 1865.
3. Accord.NET Framework v. 3.8.0, available at http://accord-framework.net

ACKNOWLEDGEMENTS

We are indebted to Devin Schweppe at Harvard University for many helpful discussions throughout the development of this work. We also gratefully acknowledge the assistance of Jimmy Eng at the University of Washington. Finally we acknowledge the important contributions of Sunandini Yedla at Thermo for assisting in testing the algorithm and workflow.

TRADEMARKS/LICENSING

PO65771-EN 0422S

ThermoFisher SCIENTIFIC