Maximizing Proteome Coverage with Advanced Peak Determination Algorithm on Tribrid Mass Spectrometers

Helene L. Cardasis[1], Graeme C. McAlister[1], Christian Thoeing[2], Philip Remeš[1], Andreas Kuehn[2], Mike Senko[1], Shannon Eliuk[1], Romain Huguet[1], Vlad Zabrouskov[1]. [1]Thermo Fisher Scientific, San Jose, CA; [2]Thermo Fisher Scientific, Bremen, Germany

ABSTRACT

Thermo Scientific™ Tribrid™ mass spectrometers are incredibly versatile instruments that combine multiple empowering MS technologies into a single platform. Working together, these technologies can sequence tens of thousands of peptides during a data-dependent LC-MS/MS analysis of a complex peptidic sample. To push sampling depths even further, we have deployed a peak determination algorithm that identifies hundreds of thousands of additional precursors. To better sample all these additional precursors, we configured the ion trap to collect MS^2 spectra at up to 40 Hz. Together these changes allow us to collect hundreds of thousands of MS^2 spectra, which translates into >35% more unique peptide identifications.

INTRODUCTION

The Thermo Scientific™ Orbitrap Fusion™ Lumos™ Tribrid™ and the Thermo Scientific™ Orbitrap Fusion™ Tribrid™ mass spectrometers typically collect ~100,000 $ITMS^2$ spectra during a two hour data-dependent LC-MS/MS analysis of a complex sample (e.g. a tryptic digest of HeLa), which converts to ~30,000 unique peptides. Though this level of proteome coverage is already extensive, off-line analysis of the same dataset by HardKlor[1] reveals hundreds of thousands of additional precursors that were never interrogated by the mass spectrometers. The majority of these unselected for MS^2 LC-MS features were never assigned charge states or monoisotopic *m/z* values by the real-time peak determination algorithm. As such, they failed to pass the typical data-dependent monoisotopic and charge state method filters. To date, every Orbitrap-equipped Thermo Scientific™ mass spectrometer has used a stripped down version of the THRASH[2-4] algorithm to assign charge states and monoisotopic *m/z* values. In the past, our variant of THRASH (aka, the "legacy peak determination" algorithm) performed well enough for typical data-dependent analyses. Now that MS instruments can collect spectra faster than 20 Hz, the shortcomings of the algorithm have become evident.

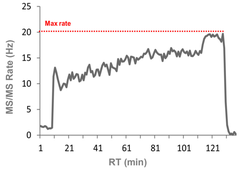


Figure 1. During a typical data-dependent LC-MS/MS analysis – including charge, monoisotopic, and dynamic exclusion precursor filtering – the average MS^2 acquisition rate is 13 Hz even though the instrument is capable of acquiring spectra at ~23 Hz.

Numerous factors can contribute to the missing or erroneous legacy algorithm peak assignments, including overlapping isotopic envelopes and poor ion statistics. Herein we demonstrate the utility of an advanced on-line Orbitrap peak determination algorithm, which overcomes many of the shortcomings of the old legacy algorithm.

MATERIALS AND METHODS

The Thermo Scientific™ Pierce™ HeLa digest protein standard was interrogated using an Orbitrap Fusion Lumos Tribrid mass spectrometer and an Orbitrap Fusion Tribrid mass spectrometer coupled to a Thermo Scientific™ Easy-nLC™ 1200 ultra-high pressure LC. We used a data-dependent (DDA) method that filtered precursors based upon charge state (2-6), monoisotopic *m/z* assignment, and dynamic exclusion (20 s). Unless noted otherwise in the text, $ITMS^2$ spectra were collected at the rapid scan rate, using an automatically determined mass range, and a maximum injection times of 35 ms. The resulting LC-MS/MS data were searched using Thermo Scientific™ Proteome Discoverer™ 2.2 software. The spectra were searched against the UniProt human database, and the peptide spectral matches were filtered to a 1% false-discovery (FDR) rate using Percolator.

The Thermo Scientific™ Pierce™ Intact Protein Standard mix was also analyzed with an Orbitrap Fusion Lumos Tribrid MS, which we coupled to a Thermo Scientific™ UltiMate™ 3000 ultra-high-pressure LC operating at 200 µL/min. We analyzed the sample with a data-dependent method that consisted of a 15,000 resolution (@200 *m/z*) Orbitrap MS^1 scan followed by data-dependent Orbitrap MS^2 scans (60,000 resolution @200 *m/z*) with the precursor charge state filter set to ≥+7.

Figure 2. All the data were collected using an Orbitrap Fusion Lumos Tribrid MS and an Orbitrap Fusion Tribrid MS.

RESULTS

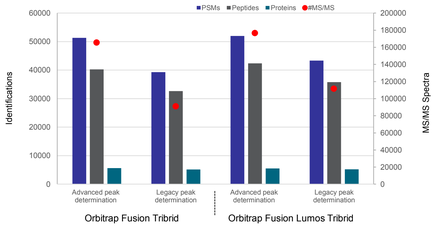
The Advanced Peak Determination (APD) Algorithm

APD algorithm boasts a suite of new features. These include the ability to annotate overlapping isotopic envelopes, improvements to the pattern matching filters used to assign the charge states and monoisotopic *m/z* values (e.g., the Patterson filter and the averagine model correlation), and a function that correlates assignments across the entire charge envelope of a given precursor (i.e., charge state deconvolution).

Between back-to-back 120 min LC-MS/MS analyses, we alternated between the legacy algorithm and the advanced peak determination algorithm.

Figure 3. In back-to-back LC-MS/MS runs we compared APD algorithm to the legacy algorithm with both Tribrids. During these analyses we used "standard" $ITMS^2$ scan settings:

- **rapid scan rate**
- **auto mass range**
- **35 ms maximum injection time**



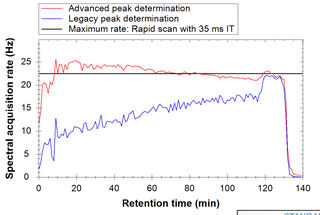


Figure 4. With the legacy algorithm used with the Orbitrap Fusion Lumos Tribrid MS, we only utilize ~60% of the $ITMS^2$ capability. With the APD algorithm, we utilize ~95%. The APD method exceeds the maximum spectral acquisition rate at the beginning of the run because the actual MS^2 injection times tend to be shorter than 35 ms and the mass range tends to be smaller than 1900 *m/z* (see Figure 6).

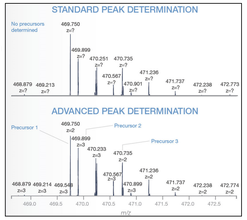
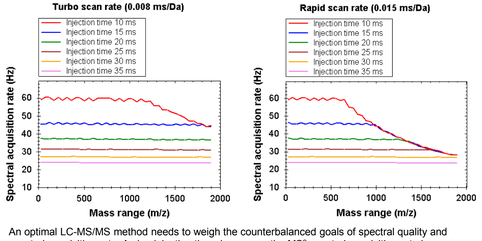


Figure 5. An example FTMS[1] mass range from the LC-MS/MS experiments where APD was *disabled* (top), and where APD was *enabled* (bottom). The mass range contains overlapping isotopic envelopes that only the APD algorithm can accurately identify. This significantly increases the population of precursors available for data-dependent analysis, which in turn results in more MS^n spectra, PSMs, and unique peptide identifications.

Optimization of the $ITMS^2$ scan settings

When the pool of available precursors was limited, it made sense to use the excess MS cycle time to collect higher quality MS/MS spectra at a slower acquisition rate. Now that we have a much larger population of precursors to interrogate, we can afford to collect more MS^2 spectra at a faster rate.

Figure 6. The $ITMS^2$ spectrum acquisition rate is determined by the analyzer scan rate, maximum injection time, and scan range. On both Tribrid MS, ion injection occurs concurrently with *m/z* analysis of the preceding scan. As such, these parameters work together to determine the maximum spectrum acquisition rate. For these graphs, ions were isolated using the quadrupole mass filter and fragmented by HCD.



An optimal LC-MS/MS method needs to weigh the counterbalanced goals of spectral quality and spectral acquisition rate. As ion injection time decreases, the MS^2 spectral acquisition rate increases but the ion statistics in the resulting MS^2 spectra decrease. Over the course of many LC-MS/MS analyses, we optimized various instrument parameters, including $ITMS^2$ maximum injection time, MS^2 mass range, MS^2 scan rate, $FTMS^1$ resolution, and MS cycle time.

Figure 7. Varying the maximum $ITMS^2$ injection during a *2 hour* LC-MS/MS analysis on the Orbitrap Fusion Lumos Tribrid MS. The scan rate was turbo and the mass range was 200-1400 *m/z*.

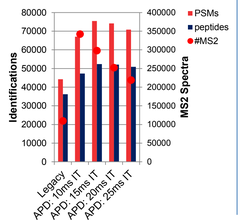
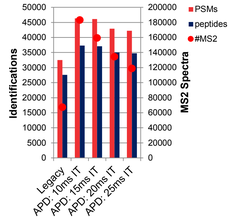


Figure 8. Varying the maximum $ITMS^2$ injection during a *1 hour* LC-MS/MS analysis on the Orbitrap Fusion Lumos Tribrid MS. The scan rate was turbo and the mass range was 200-1400 *m/z*.



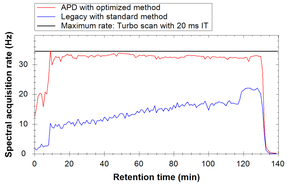


Figure 9. For a 2 hour LC-MS/MS method, the optimal maximum injection time is 20 ms with the Orbitrap Fusion Lumos Tribrid MS. Accounting for scan time overhead, and the MS time spent collecting MS^1 spectra, the ion trap can acquire spectra at a minimum rate of ~35 Hz when the ion injection time is ≤ 20 ms. With APD enabled, the Orbitrap Fusion Lumos Tribrid MS uses ~93% of this capacity.

Replicate Analyses Using APD LC-MS/MS with Both Tribrid Mass Spectrometers

For LC-MS/MS methods with APD enabled, we found that the following $ITMS^2$ settings nicely balanced sensitivity and versatility: 20 ms maximum injection time, rapid scan rate, and an automatic mass range. Using these settings, we performed replicate (n≥3) LC-MS/MS analyses comparing APD on vs. APD off. For both conditions, we injected 1 µg of a tryptic HeLa digest, and we performed both 1 and 2 hour LC gradients. During the legacy-based LC-MS/MS method, we used our "standard" $ITMS^2$ scan settings: 35 ms maximum injection time, rapid scan rate, and an automatic mass range.

Figure 10. During a 2 hr LC-MS/MS method we identified >35% more unique peptides with APD. Or as an alternative, with APD enabled we identified an equivalent number of unique peptides with an LC-MS/MS method that is half as long as the 2 hr legacy-based LC-MS/MS method.

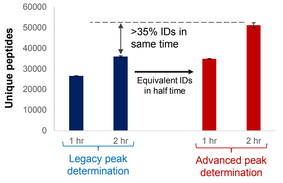


Figure 11. We compared the overlap between 3 x APD 2 hr LC-MS/MS analyses to the overlap between 3 x legacy-based 2 hr LC-MS/MS analyses. The number of unique peptides observed in all three replicates increased by ~50% with APD (26,158 vs. 38,045).

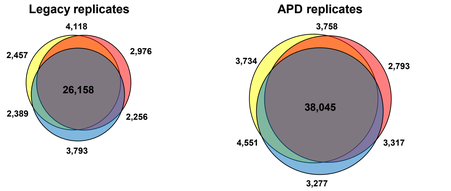


Figure 12. The different populations in the Venn diagrams above (peptides seen in 1, 2, or all 3 replicates) were distributed on a histogram by precursor intensity. In this figure the three populations are stacked (i.e., replicate 1 = replicate 1, replicate2 = 1+2, and replicate 3 = 1+2+3), such that the entire histogram represents all the peptides identified in all the replicates.

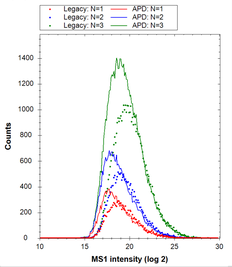
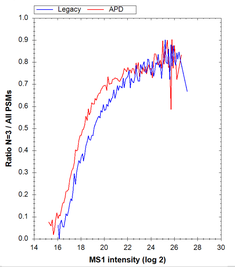


Figure 13. Using the distributions in Figure 12, at each intensity bin we calculated the ratio between the number of peptides identified in all three replicates over the total number of peptides. Based upon this analysis, the limit of reproducible detection (where > 50% of all the peptides identified were observed in all three replicates) is ~2-3x lower with APD.



TOP-DOWN RESULTS

Demonstration of the Utility of APD for Top-Down Analysis

The APD algorithm's ability to identify overlapping isotopic envelopes is the main force behind most of the gains we observed during the LC-MS/MS analyses of peptide samples. However, the APD algorithm also has improved charge state assignment functions, including the ability to correlate assignments across the entire precursor charge envelope. These other APD advancements greatly improve large biomolecule charge state assignment.

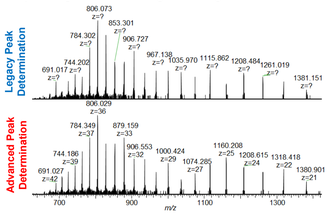


Figure 14. We analyzed the Pierce intact protein standard mix by LC-MS/MS. FTMS[1] spectra were collected at 15k and were the summation of 5 uscans. In back-to-back runs, we tested APD *on* and *off*. With APD *on*, the instrument can accurately identify highly charged and complex protein envelopes. Thanks to the improved charge state assignment abilities of the APD algorithm, we can execute more advanced top-down methods – such as, MS/MS fragmentation of one charge state per precursor.

CONCLUSIONS

- The advanced peak determination (APD) algorithm identifies hundreds of thousands of additional precursors in Orbitrap spectra for data-dependent analysis.
- We observe a large increase in the unique peptide identifications when we configure the quadrupole ion trap to favor a faster MS/MS rate that more effectively samples all the additional precursors.
- With APD and the optimized $ITMS^2$ method, we collect ~250,000 MS^2 spectra during a 2 hr LC-MS/MS method. This converts into >45,000 unique peptide IDs, which is a >35% improvement over the conventional legacy algorithm based approach.
- This improved sample coverage translates into better run-to-run reproducibility. Such that, the limit of reproducible identification is ~2-3x lower with APD.
- APD also enables improved large molecule charge state assignment and in turn top-down data-dependent decisions.

For Research Use Only. Not for use in diagnostic procedures.

REFERENCES

1. Hoopmann MR, Finney GL, MacCoss MJ. High speed data reduction, feature selection, and MS/MS spectrum quality assessment of shotgun proteomics datasets using high resolution mass spectrometry. *Anal. Chem.* **2007**, 79, 5630-5632.
2. Senko MW, Beu SC, McLafferty FW. Automated assignment of charge states from resolved isotopic peaks for multiply charged ions. *J. Am. Soc. Mass Spectrom.* **1995,** 6, 52-56.
3. Senko MW, Beu SC, McLafferty FW. Determination of monoisotopic masses and ion populations for large biomolecules from resolved isotopic distributions. *J. Am. Soc. Mass Spectrom.* **1995,** 6, 229-233.
4. Horn DM, Zubarev RA, McLafferty FW. Automated reduction and interpretation of high resolution electrospray mass spectra of large molecules. *J. Am. Soc. Mass Spectrom.* **2000,** 11, 320-332.

ACKNOWLEDGEMENTS

The authors would like to thank Michael Westphall, Jae Schwartz, John Syka, Jesse Canterbury, Chris Mullen for helpful conversations.

TRADEMARKS/LICENSING

ThermoFisher SCIENTIFIC

PO65303-EN 0518S