Maximizing proteome coverage through improved on-line Orbitrap peak determination

Graeme C. McAlister[1], Christian Thoeing[2], Helene L. Cardasis[1], Alex S. Hebert[3], Romain Huguet[1], Philip Remeš[1], Andreas Kuehn[2], Mike Senko[1], Shannon Eliuk[1], Joshua J. Coon[3], Vlad Zabrouskov[1];
[1]Thermo Fisher Scientific, San Jose, CA; [2]Thermo Fisher Scientific, Bremen, Germany; [3]University of Wisconsin, Madison, WI

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 new peak determination algorithm that identifies hundreds of thousands of additional precursors. To better sample all these additional precursors, we optimized the ion trap MS2 scan settings. Together these changes allow us to collect hundreds of thousands of MS2 spectra, which translates into >35% more unique peptide identifications.

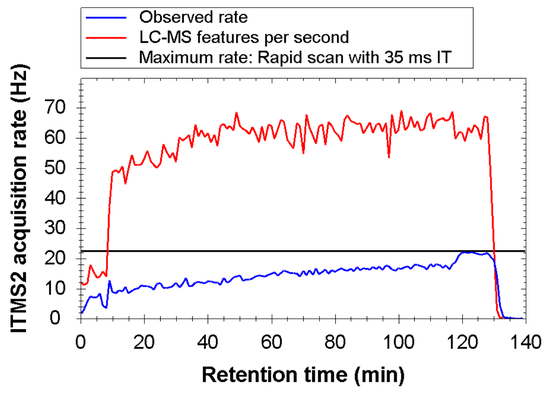
INTRODUCTION

The Thermo Scientific™ Orbitrap Fusion Lumos™ Tribrid™ mass spectrometer typically collects ~100,000 ITMS2 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.

The majority of these un-fragmented 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. But now that MS instruments can routinely collect spectra faster than 20 Hz, the shortcomings of the legacy algorithm have become evident.

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.

Figure 1. During a typical data-dependent LC-MS/MS analysis – including charge, monoisotopic, and dynamic exclusion precursor filtering – the average MS2 acquisition rate is 13 Hz even though the instrument is capable of acquiring spectra at ~23 Hz.



MATERIALS AND METHODS

The Pierce™ HeLa digest protein standard was interrogated using an Orbitrap Fusion Lumos™ Tribrid™ mass spectrometer coupled to a Thermo Scientific™ Easy-nLC™ 1000 ultra-high pressure LC. We used a data-dependent method that filtered precursors based upon charge state (2-6), monoisotopic m/z assignment, and dynamic exclusion (20 sec). Unless noted otherwise in the text, ITMS2 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™ 2.1 software. The spectra were searched against the UniProt human database, and the peptide spectral matches were filtered to a 1% false-discovery rate using Percolator.

The Thermo Scientific™ Pierce™ Intact Protein Standard mix was also analyzed with an Orbitrap Fusion Lumos Tribrid mass spectrometer, which we coupled to a Thermo Scientific™ Dionex™ UltiMate™ 3000 ultra-high pressure LC operating at 200ul/min. We analyzed the sample with a data-dependent method that consisted of a 15k resolution (@200m/z) Orbitrap MS1 scan followed by data-dependent Orbitrap MS2 scans with the precursor charge state filter set to ≥+7.

RESULTS

Initial characterization of Advanced Peak Determination algorithm

The Advanced Peak Determination (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). To assess APD performance, we alternated between the legacy algorithm and the advanced peak determination algorithm in back-to-back LC-MS/MS analyses .

Figure 2. Example FTMS1 spectra 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.

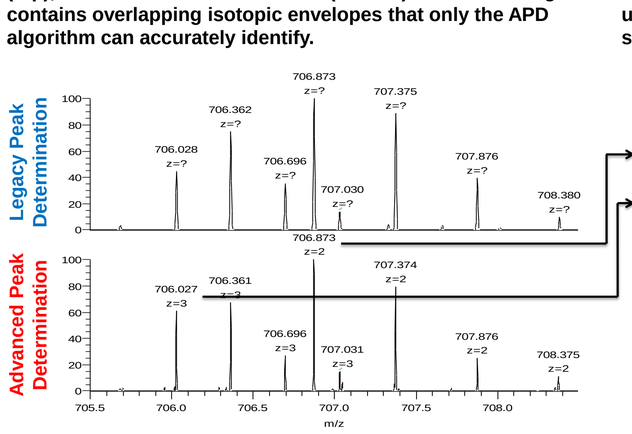


Table 1. The precursors in the mass range (Fig. 2) were both identified by Proteome Discoverer. Both of the MS2 spectra used to generate these PSMs were collected during the same MS cycle of the APD LC-MS/MS analysis.

	Mono m/z	Charge	Sequence	XCorr
Precursor #1	706.873	2	EPALNEANLS NLK	3.385
Precursor #2	706.027	3	ASLLQNESTN EQLQIHYK	3.568

Figure 3. In back-to-back LC-MS/MS runs we compared the APD algorithm to the legacy algorithm. For these analyses we used "standard" MS2 settings: rapid scan rate, auto mass range, and 35 ms maximum injection time.

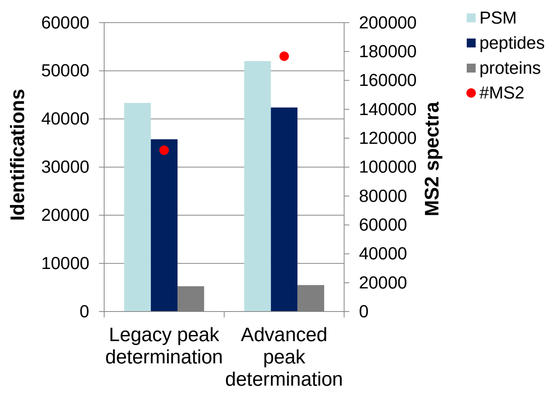
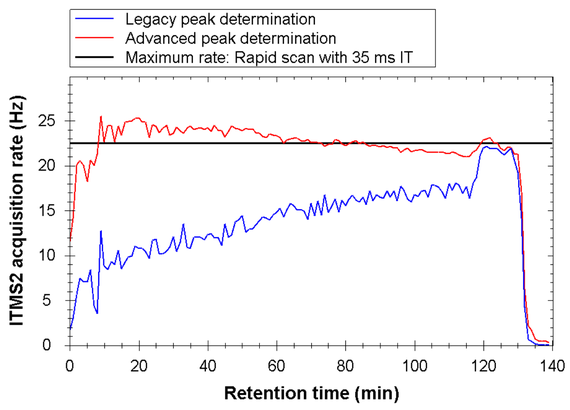


Figure 4. The legacy algorithm utilizes ~60% of the ITMS2 capacity, while APD utilizes ~95%. The APD method exceeds the "maximum" acquisition rate at the beginning of the run when the average MS2 injection times are less than 35 ms and the mass range is less than 1900 m/z (Fig. 5).



Optimization of the ITMS2 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 slower acquisition rates. Now that we have a much larger population of precursors to interrogate, we can afford to collect more MS2 spectra at a faster rate.

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 MS2 spectral acquisition rate increases but the ion statistics in the resulting MS2 spectra decrease. Over the course of many LC-MS/MS analyses, we optimized various instrument parameters, including ITMS2 maximum injection time, MS2 mass range, MS2 scan rate, FTMS1 resolution, and MS cycle time.

Figure 5. The ITMS2 spectral acquisition rate is determined by injection time and scan range. On the Tribrid MS, ion injection occurs concurrently with m/z analysis of the preceding scan. For these graphs, ions were isolated in the quadrupole mass filter and fragmented by HCD.

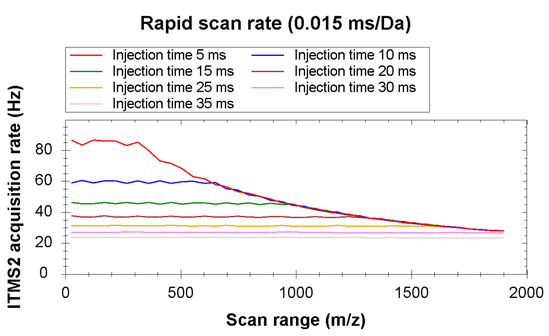
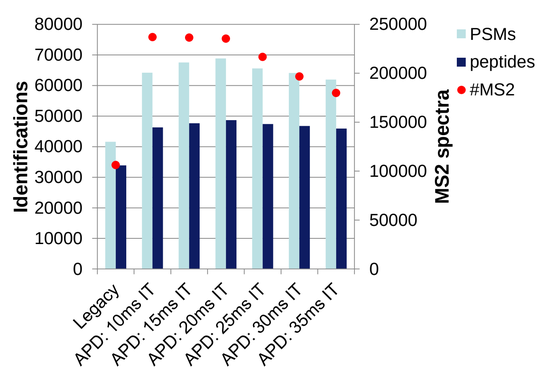


Figure 6. Varying the maximum ITMS2 injection during a 2 hour LC-MS/MS analysis. The scan rate was turbo and the mass range was 200-1400 m/z.



Replicate analyses using APD LC-MS/MS

For LC-MS/MS methods with APD enabled, we found that the following ITMS2 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 injected1 ug 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" ITMS2 scan settings: 35 ms maximum injection time, rapid scan rate, and an automatic mass range.

Figure 7. During a 2 hr LC-MS/MS method we identified >35% more unique peptides with APD. Or as an alternative , with APD we identified an equivalent number of peptides in half the time it would have taken with the legacy algorithm.

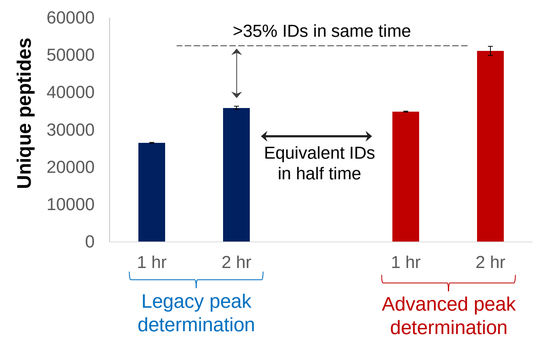


Figure 8. We compared the overlap between 3 x APD 2 hr LC-MS/MS analyses to the overlap between 3 x Legacy-based 2 hr LC-MS/MS analyses. The number of unique peptides observed in all three replicates increased by ~50% with APD (26,158 vs. 38,045).

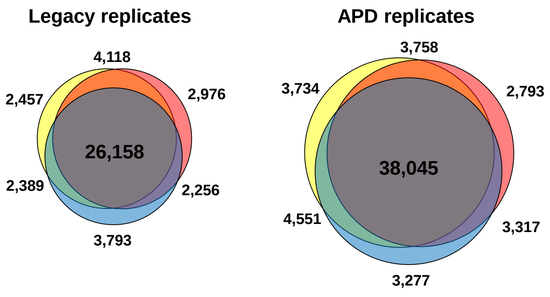


Figure 9. The different populations in the Venn diagrams in figure 8 (peptides seen in 1, 2, or all 3 replicates) were distributed on a histogram by precursor intensity. In this figure the three populations are stacked (i.e., replicate 1 = replicate 1, replicate 2 = 1+2, and replicate 3 = 1+2+3).

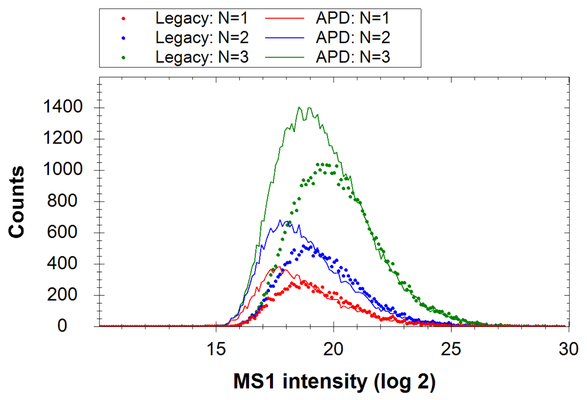
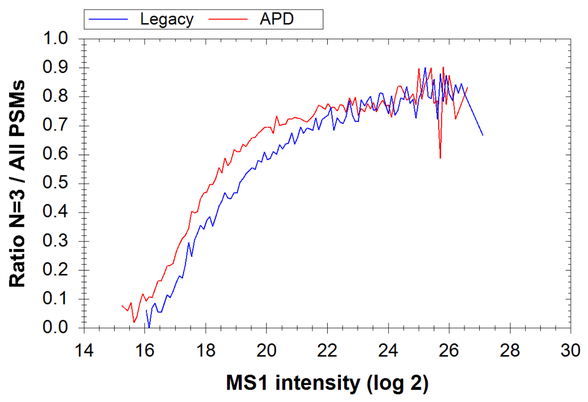


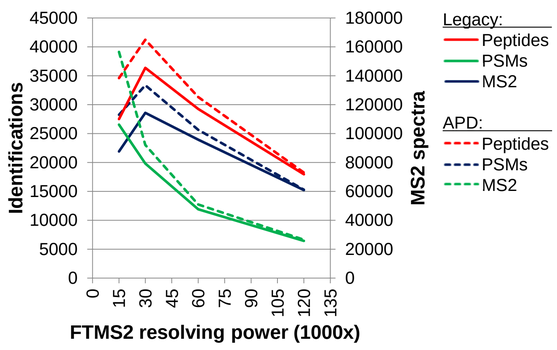
Figure 10. Using the distributions in figure 9, at each intensity bin we calculated the ratio between the number of peptides identified in all three replicates over the total number of peptides. Based upon this analysis, the limit of reproducible detection is ~2-3x lower with APD.



Examining the relationship between APD and spectral acquisition rate

The benefits of APD are tied closely to the MS2 spectral acquisition rate. At short FTMS2 transients (32 ms or 15k RP), we collect FTMS2 spectra at >20 Hz and we observe a ~30% increase in the number of unique peptides identified during the APD run. At longer FTMS2 transients (256 ms or 120k) we observe no significant improvement with APD. As the FTMS2 resolving power increases, the MS2 spectral acquisition rate decreases. As such, it takes fewer and fewer precursors to utilize the full MS2 sampling capacity of the method. Or in other words, at these slow MS2 interrogation rates, there are enough precursors identified by both the APD and SPD algorithms to keep the instrument busy during the entire LC-MS/MS analysis.

Figure 11. We measured the number of FTMS2 spectra collected, peptide spectral matches, and unique peptides, as a function of MS2 resolving power and APD status.



Demonstration of the utility of APD for top-down analysis

The ability of the APD algorithm to identify overlapping isotopic envelopes is the driving 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.

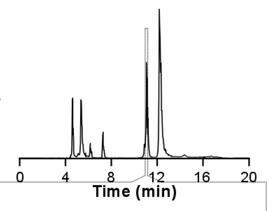
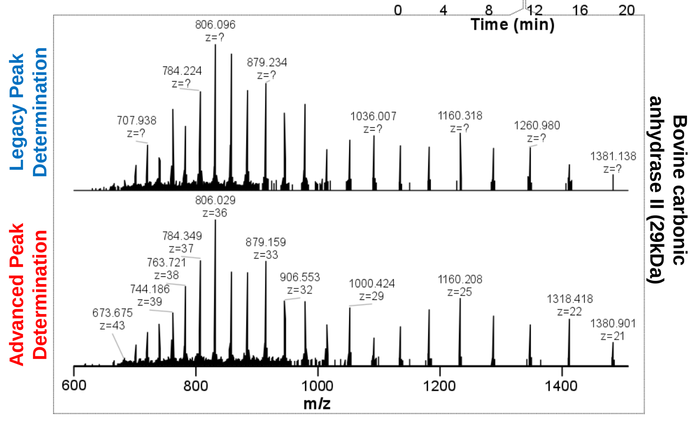


Figure 12. We analyzed the Pierce intact protein standard mix by LC-MS/MS. FTMS1 spectra were collected at 15k and were the summation of 5 uscans. In back-to-back runs we tested APD *on* and *off*. With APD *on*, the instrument can accurately identify highly charged and complex protein envelopes. Thanks 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 ITMS2 method, we collect ~250,000 MS2 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 has improved large molecule charge state assignment and in turn this enables more intelligent top-down data-dependent methods.

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, and Rosa Viner for helpful conversations.

TRADEMARKS/LICENSING