Published in final edited form as:

Anal Chem. 2018 February 06; 90(3): 2333–2340. doi:10.1021/acs.analchem.7b04808.

Improved Precursor Characterization for Data-Dependent Mass Spectrometry

Alexander S. Hebert¹, Christian Thöing⁴, Nicholas M. Riley², Nicholas W Kwiecien¹, Evgenia Shiskova³, Romain Huguet⁵, Helene L. Cardasis⁵, Andreas Kuehn⁴, Shannon Eliuk⁵, Vlad Zabrouskov⁵, Michael S. Westphall¹, Graeme C. McAlister⁵, and Joshua J. Coon^{1,2,3,6,*}

¹Genome Center of Wisconsin, University of Wisconsin-Madison, Madison, WI, 53706, USA

²Department of Chemistry, University of Wisconsin-Madison, Madison, WI, 53706, USA

³Department of Biomolecular Chemistry, University of Wisconsin-Madison, Madison, WI, 53706, USA

⁴Thermo Fisher Scientific, 28199, Bremen, Germany

⁵Thermo Fisher Scientific, San Jose, CA 95134, US

⁶Morgridge Institute for Research, Madison, WI, 53406 USA

Abstract

Modern ion trap mass spectrometers are capable of collecting up to 60 tandem MS (MS/MS) scans per second, in theory providing acquisition speeds that can sample every eluting peptide precursor presented to the MS system. In practice, however, the precursor sampling capacity enabled by these ultra-fast acquisition rates is often underutilized due to a host of reasons (e.g., long injection times and wide analyzer mass ranges). One often overlooked reason for this underutilization is that the instrument exhausts all the peptide features it identifies as suitable for MS/MS fragmentation. Highly abundant features can prevent annotation of lower abundance precursor ions that occupy similar mass-to-charge (m/z) space, which ultimately inhibits the acquisition of an MS/MS event. Here we present an advanced peak determination (APD) algorithm that uses an iterative approach to annotate densely populated m/z regions to increase the number of peptides sampled during data-dependent LC-MS/MS analyses. The APD algorithm enables nearly full utilization of the sampling capacity of a quadrupole-Orbitrap-linear ion trap MS system, which yields up to a 40% increase in unique peptide identifications from whole cell HeLa lysates (approximately 53,000 in a 90-minute LC-MS/MS analysis). The APD algorithm maintains improved peptide and protein identifications across several modes of proteomic data acquisition, including varying gradient lengths, different degrees of pre-fractionation, peptides derived from multiple proteases, and phosphoproteomic analyses. Additionally, the use of APD increases the number of peptides

*Corresponding author: jcoon@chem.wisc.edu.

ASSOCIATED CONTENT

Supporting Information Available: Supplemental figures and expanded materials and methods. This material is available free of charge via the Internet at <http://pubs.acs.org>.

CT, RH, HLC, AK, SE, VZ, and GCM are employees at Thermo Fisher Scientific, which develops and distributes instruments equipped with, and upgrades for, the APD algorithm. Raw data are available on Chorus (Project ID 1415)

characterized per protein, providing improved protein quantification. In all, the APD algorithm increases the number of detectable peptide features, which maximizes utilization of the high MS/MS capacities and significantly improves sampling depth and identifications in proteomic experiments.

Graphical abstract

INTRODUCTION

Over the last several decades mass analyzer performance has greatly improved. Years ago, low resolution MS^1 scans limited the number of features that could be confidently selected for MS/MS. No matter, MS/MS scan rates were so low that only the most abundant precursors could be sampled (*i.e.*, top three or five).^{1,2} Beginning in the late 1990s time-of-flight mass analyzers began to offer improved MS^1 mass resolution.^{3–5} Hybrid quadrupole linear ion trap Fourier transform ion cyclotron resonance mass spectrometers became commercially available shortly after that and offered up to 100,000 mass resolving power.⁶ With these devices, MS^1 spectra rich with features were routinely produced, yet MS/MS scan rates still lagged at ~ 3 Hz. In 2006 Orbitrap hybrids became available and advanced rapidly over the past decade;^{7–13} meanwhile, scan rates have steadily increased so that today MS/MS scans can be collected at a staggering 60 Hz (Figure S1). With such a system, a complex mixture of tryptic yeast peptides analyzed over a 70 minute separation produced $\sim 200,000$ unique LC- MS^1 features – a close match for the number of MS/MS scans that can be acquired in that time at 60 Hz ($\sim 252,000$). In actuality, only 31% (65,511) of these detected features were sampled by MS/MS.

Tandem mass spectrometry (MS/MS) is the central technology for proteome analysis.^{14–18} Following their elution from a chromatography column, and conversion to the gas-phase, peptide cations have their mass-to-charge ratios (m/z) measured in an MS^1 survey scan. This MS^1 scan is leveraged in two ways. Peptide cations are detected as distinct clusters of m/z peaks corresponding to their charge state and isotopic distribution. An MS^1 scan typically contains evidence for dozens, sometimes even hundreds, of unique peptide ion clusters. These MS^1 signals, *i.e.*, peptide features, are detected using embedded algorithms. During data-dependent MS/MS methods, the instrument identifies and prioritizes these detected features for subsequent MS/MS analysis. Note only those features selected for MS/MS have an opportunity to be identified. The MS^1 scan also reveals each precursor's intact mass –

data that when combined together with the MS/MS scan can allow identification of the precursor peptide sequence.

Advances in throughput and achievable depth in proteomic experiments are often catalyzed by improvements in instrumentation. Such examples include increased sensitivity, faster scanning, and more intelligent control software to increase the number of peptides sequenced per unit time.^{12,19,20} When analyzing whole proteomes, hundreds of thousands of peptide features are presented to the mass spectrometer over the course of an LC-MS/MS gradient.²¹ Ideally the instrument would select each feature for MS/MS allowing for sampling of all detected features. In reality, the mass spectrometer can only sample a fraction of the features it detects for MS/MS fragmentation due to restrictions in scan speed and the effective dynamic range of the mass analyzer. Over the past decade, as instruments with improved MS/MS acquisition rates have been introduced, proteomic sampling depth concomitantly increased. Mann and co-workers estimated in 2011 that an MS/MS rate of 25 Hz would be needed to sample all peptide features in an analysis of a complex proteome, and, indeed, the introduction of an instrument with an MS/MS rate of ~20 Hz enabled the characterization of the yeast proteome in just over an hour.²¹

The Orbitrap tribrid mass spectrometer is capable of collecting up to 60 ion trap MS/MS scans per second; however, the increase in proteomic depth achievable with this instrument has not linearly matched the improvement in MS/MS acquisition speed. We recently showed that the high acquisition rates of modern instruments are significantly underutilized in most experiments because the instrument exhausts the detectable number of peptide features to select for MS/MS from a given survey scan.²² One avenue to alleviate this problem is to improve chromatographic separations so that peptides do not co-elute, decreasing MS¹ peak complexity, although we are not aware of a chromatographic implementation that could achieve this goal.²³

An alternative approach to better utilizing the high MS/MS capacity of modern instruments is to improve the ability of the instrument to determine which features are peptide precursors, suitable for MS/MS fragmentation. Criteria for selecting an isotope distribution for subsequent MS/MS selection often includes charge state (e.g., +2) and monoisotopic m/z assignment filtering (which acts as a proxy for whether an isotopic distribution resembles a typical peptidic feature). The current standard peak determination (SPD) algorithm (i.e., the Q Exactive and Orbitrap Fusion Tribrid) only considers local maxima, in a given mass-to-charge (m/z) region, when annotating the most abundant peptide feature, and is unable to annotate additional potential peptides in the immediate area.

Here we describe a new peak annotation algorithm, called advanced peak determination (APD) that can significantly improve the number of sampled peptide features in a given experiment by using a fast, iterative approach to annotate several isotope distributions that occupy an m/z region. The APD algorithm increases the number of unique peptides identified by up to 40% over standard peak determination, equating to approximately 53,000 unique tryptic peptides identified from a whole cell HeLa lysate in a 90 minute LC-MS/MS analysis. This new peak annotation approach also increases unique peptide identifications for three other proteases (chymotrypsin, gluC, and lysC) and maintains improved proteome

characterization over SPD when pre-fractionation is used to simplify the proteome digest across multiple experiments. In all, the APD algorithm enables near complete utilization of the rapid MS/MS acquisition rates of the newest Orbitrap tribrid instrument and represents a direct approach to improving proteomic sampling.

MATERIALS AND METHODS

LC-MS/MS

Digestion conditions are described in the supporting information. HeLa tryptic peptides, and mouse brain GluC, chymotrypsin, or lysC peptides, were each dissolved in 0.2% formic acid at concentration of 1 $\mu\text{g}/\mu\text{L}$, mouse brain phosphopeptides were dissolved in 40 μL 0.2% formic acid. All analyses were performed with 1 μg of peptides injected on column, except the phosphorylation analyses, where 10% of total material was injected each time. Separations were performed over in-house fabricated 70 μm inner diameter $\times$ 360 μm outer diameter columns with an integrated nano electrospray emitter packed 30 cm long with 1.7 μm C18 bridged ethylene hybrid particles (Waters, Milford, MA).¹³ Separation times noted in the text include loading time in 100% A (0.2% formic acid), gradient elution in increasing % B (0.2% formic acid/70% acetonitrile), and re-equilibration times in 100% A. An additional overhead of ~3 minutes for loading the sample into the sample loop is not included. All separations were performed with a Thermo Dionex Ultimate 3000 RSLC-nano liquid chromatography instrument and an in house fabricated column heater to conduct separations at 50 °C.

Eluted peptides were analyzed on an Orbitrap Fusion Lumos Tribrid platform with Instrument Control Software version 3.0. Typical analyses utilized a 240,000 resolving power survey scan with an AGC of 10^6 , followed by MS/MS of the most intense precursors for 1 second. The MS/MS analyses were performed by 0.7 m/z isolation with the quadrupole, normalized HCD collision energy of 30%, and analysis of fragment ions in the ion trap using the “Rapid” scan speed scanning from 200–1,200 m/z . Dynamic exclusion was set to 20 seconds, Monoisotopic Precursor Selection (MIPS) was set to Peptide. For MS/MS analyses the maximum injection time was set to 18 msec, with an AGC target of 30,000, and charge states unknown, 1, or 5 were excluded. The advanced peak determination was either toggled off or on for SPD and APD analyses, respectively. Phosphoproteomic experiments were performed similarly except for FTMS² analyses were performed with 60,000 resolving power survey scan. MS/MS scans were analyzed with 30,000 resolving power, AGC target set to 200,000, and 1.6 Da isolation window. Phosphoproteomic analyses were each 180 minutes long.

Data for the FTMS² experiments were collected as described above with the following exceptions. Survey scans were collected at 60,000 resolving power and AGC set to 10^6 , the MS/MS AGC target was set to 50,000, and dynamic exclusion was 30 seconds. The maximum MS/MS injection time is dictated by the transient length and set to the recommended value (6, 22, 54, 86, 118, and 246 msec was used for the 7,500, 15,000, 30,000, 50,000, 60,000, and 120,000 resolving power scans, respectively). These analyses were each 90 minutes long, including loading, wash, and equilibration times. See supporting information for more detail.

RESULTS AND DISCUSSION

To identify and prioritize MS¹ features for MS/MS analysis, the standard peak determination algorithm, traditionally employed on all Orbitrap hybrid MS systems, scans across m/z space, finds local maxima, and attempts to assign charge states based on m/z differences between isotopes.²⁴ For each isotopic cluster that is assigned a charge state, the monoisotopic peak is annotated by comparing the isotopic intensity distribution against that of a theoretical peptide composed of averagine with the same mass.²⁵ Following detection of a precursor feature, the algorithm moves up two m/z units to scan for the next local maximum. When multiple features have isotopic clusters that are within two m/z units of one another, only the most intense feature is considered for charge state and monoisotopic m/z value annotation. As such, the lower abundance features are not selected for data-dependent MS/MS if the method requires charge state and monoisotopic m/z assignment. We suppose that when initially developed, this relatively simple approach worked quite well as the number of accessible features vastly outnumbered the available MS/MS capacity. But today, MS/MS scan rates have increased so that sampling capacity has outrun the standard algorithm's ability to identify quality targets. We conclude that design and implementation of a new fast and intelligent peak determination algorithm could substantially boost performance of data-dependent acquisition methods, especially on newer, fast-scanning MS systems.

Here we describe an advanced peak determination (APD) algorithm optimized for computational speed on modern multi-core systems. The algorithm was predominantly developed for peptide and protein analysis; however, it can, in principle, be applied to other analyte classes. Figure 1 presents the APD logic. First, the MS¹ spectrum, a list of centroid peaks, is divided into multiple m/z bins that are evenly distributed to the available processor cores. In this way, a multithreading environment is created where each core has an equivalent share of the workload to process. Second, APD extracts distinct isotope distributions. Starting with the most intense peak within an m/z bin and using m/z distances of adjacent peaks, APD calculates and scores a range of potential precursor charge states using a set of metrics that leverage the observed intensity patterns and m/z deviations. Once a charge state is assigned, m/z peaks belonging to this isotope distribution are annotated and removed from the peak list. Note these m/z peaks still exist in the spectrum and may be assigned to multiple isotope distributions. Distinct isotope distributions are extracted from the m/z bin in an iterative manner, in descending intensity order, until the peak list is exhausted. Averagine 'fits' are used to assign monoisotopic masses. The third, and final, step is charge state deconvolution, which is particularly helpful for higher charged precursors (*e.g.*, intact proteins). Here the algorithm attempts to correlate individual isotope distributions to the same analyte (*e.g.*, a protein's entire charge state envelope).

The APD algorithm annotates more MS¹ isotope distributions, especially in m/z regions of high complexity. Figure 2 provides a comparison of how the standard and APD algorithms annotate complex isotope distributions of overlapping peptide features in a three m/z window. Here the standard algorithm annotates only the most intense precursor (green, $z = 3$). The second panel of Figure 2 shows the same m/z regions just a few moments later in the separation, here the standard algorithm identifies the most abundant feature again, which is

now a doubly protonated precursor at higher m/z (blue). The lowest m/z species in the cluster (yellow) was never annotated by the standard algorithm. This exemplifies that the standard algorithm cannot annotate overlapping peaks in a given spectrum. Alternatively, APD correctly assigns all three precursors in both MS¹ scans. The net result is a reproducible, marked improvement in the assignment of overlapping features with APD.

Parallelization of MS¹ transient acquisition with precursor isolation, fragmentation, and MS/MS analysis enables the quadrupole-Orbitrap-linear ion trap tribrid to have MS/MS acquisition rates upwards of 60 Hz (Figure 1 S1), while still affording good spectral quality. Several factors, such as speed, signal, and resolution, contribute when determining an optimal MS/MS scan rate. Knowing 60 Hz may not always give the highest quality data, we opted to collect MS/MS spectra at 35–40 Hz (~ 2,100 MS/MS scans per minute). We tested performance of the standard algorithm to detect features, and trigger MS/MS scans, during a one hour separation of tryptic peptides from HeLa cells. Figure 3a demonstrates that the standard algorithm does not identify sufficient peptide features to fully utilize the 35 Hz MS/MS capacity (72%). Decreasing the dynamic exclusion time could increase the scan frequency; however, such a strategy only increases repetitive sampling of the most abundant features and offers limited improvements to proteomic coverage. Analysis of the same sample with the APD algorithm, however, permitted near full utilization (96%) of the MS/MS capacity (Figure 3b).

The APD algorithm increases annotation of more MS¹ features in m/z regions of high complexity. We compared the peptide features detected, selected for MS/MS, and identified between the standard and APD algorithms across the entire LC-MS/MS experiment (Figure 4a). Note the number of available peptide features was determined using a post-acquisition data analysis software (MaxQuant). These data demonstrate that APD provides a considerable improvement in sampled features with the greatest benefit from m/z regions of the highest feature density (400–900 m/z). Since APD performs peak determination of overlapping peptide features, we reasoned that the additional MS/MS scans might result from co-isolation of multiple precursors. To evaluate we calculated precursor purity for all MS/MS scans in both the standard and APD algorithm datasets. Not surprisingly, features selected by the standard algorithm were on average 71% pure, whilst those from the APD method were 56% pure (Figure 4b). This observation is further corroborated by a reduction in the identification success rate (identifications/total MS/MS spectra acquired), from ~50% with the standard algorithm to ~40% with APD. Reduced identification rates of spectra occur when precursors have elevated interfering product ions from nearby contaminants (Figure 4b). Note that this reduction in the success rate is easily offset during these experiments by an increase in the total number of MS/MS acquired. The increased sampling of lower purity precursors can have other impacts on the experiment, besides reduced success rate.

The standard algorithm can act as a crude precursor purity filter during a data-dependent method. As detailed above, the standard algorithm does not assign charge states or monoisotopic m/z values if the precursor purity is too low. For methods with very slow MS/MS spectral acquisition rates, the standard algorithm will rebuff precursors of low purity and high intensity in favor of higher purity but lower intensity precursors (Figure 4c). If there

are enough precursors available to utilize the full MS/MS sampling capacity, then it's beneficial to only go after the precursors with the highest purity. Methods with very slow spectral acquisition rates can in some specific cases cause the standard algorithm to outperform APD (see phosphorylation FT example below). In the context of methods with moderate to fast spectral acquisition rates (*e.g.*, the ion trap MS/MS methods), APD is clearly beneficial as it utilizes the full MS/MS sampling capacity of the method and instrument (Figure 3). Or in other words, when the method is capable of scanning fast, then the best option is to simply sample as much as possible.

Having confirmed the APD algorithm's ability to detect more features, we sought to optimize and assess APD's large-scale performance. First, we considered the impact of MS¹ scan resolution, where there often exists a trade-off between speed and sensitivity. Higher resolving power MS¹ scans take longer, thereby decreasing scan rate; however, the boosted resolution can allow detection of more features. The time penalty incurred by higher resolution MS¹ scans can be mitigated in some instrument configurations by parallelizing MS/MS and MS¹ scans. This strategy, however, only works so long as the MS/MS analyzer is working during the entire MS¹ acquisition period. Figure 5a explores these complex relationships for both the standard and APD algorithms. When using standard algorithm, as the MS¹ resolution is increased, the MS/MS scan rate significantly decreases as there are not sufficient MS/MS scans to utilize the longer Orbitrap analysis time. In contrast, the APD algorithm continues to annotate more precursors with increasing MS¹ resolution so that the most MS/MS scans are triggered at the highest resolving power setting (500,000). Using optimal settings (240,000 resolving power) many more MS/MS events are triggered, and the APD algorithm produces ~20% more unique peptide identifications during a one hour analysis (40,000 vs. 33,000). We also note that the accuracy of the monoisotopic isotope is modestly improved (from 92% to 95% accurate) with the APD algorithm (Figure S2).

Next, we sought to determine whether the benefits of APD would persist over varying analysis conditions and sample types. First, we determined the number of unique peptides that could be identified with varied single shot LC-MS/MS gradients – ranging from 60 to 180 minutes. Figure 5c reveals no matter the gradient duration, the APD algorithm solidly outperforms the standard algorithm. Figure 5d presents the number of protein groups that are identified as a function of sample pre-fractionation (*i.e.*, two-dimensional LC-MS/MS) for the two algorithms. Again, APD always outperforms and, following analysis of just eight fractions, nearly 9,000 proteins were identified. Additionally, the increased proteomic coverage correlated with improved LFQ quantitation from injection replicates. Protein measurements with <10% RSD increased by 7% with APD, while measurements with >20% RSD decreased by 13% (Figure S3a). This is likely attributed to the overall increase in peptides per protein in the analyses using APD (Figure S3).

Next we analyzed complex mixtures of peptides generated by the proteases chymotrypsin, GluC, and LysC from a mouse brain sample (separately, Figure 6a). In each case, analyses using APD allowed identification of ~ 20% more unique peptides as compared to the standard algorithm. Finally, we analyzed complex mixtures of tryptic phosphopeptides and, at the high MS/MS acquisition rates of the ion trap, the APD method provided similar gains over the standard approach (Figure 6b). Analyses using APD provided slightly fewer

phosphopeptide identifications in the analyses using the Orbitrap for MS/MS detection, likely due to a combination of factors including the decreased parallelization of Orbitrap/Orbitrap acquisition strategies and potentially due to the decreased precursor purities (*vide supra*). Additionally, we found that ion trap collisional activation performance was similarly improved by APD but to a less extent than HCD, likely due to decreased spectral acquisition rate as the collisional activation requires approximately 10 ms (Figure S4).

Finally, we evaluated the benefits of using APD for FT MS/MS analysis at varied spectral acquisition rates. In Figure 7, we measured the number of MS/MS spectra, and unique peptides detected as a function of FT MS/MS resolving power. Concurrently, we varied the maximum precursor ion injection time in proportion to the Orbitrap transient length. We decreased the maximum injection time as the transient length decreased, such that the precursor ion injection time was never longer than the transient. Due to these varied FT MS/MS settings, the spectral quality increases significantly from the shortest transient to the longest. At the shortest Orbitrap transient tested (16 ms or 7,500 resolving power) we observed a ~25% increase in the number of unique peptides identified in the analysis using APD, while at the longest transient tested (256 ms or 120K Resolution) we observed little difference between the APD and standard algorithms. As the FT MS/MS resolving power increases, the MS/MS spectral acquisition rate decreases. As such, it takes fewer and fewer precursors to utilize the full MS/MS sampling capacity. Or in other words, at these slow MS/MS interrogation rates, there are enough precursors identified by both algorithms to keep the instrument busy.

CONCLUSION

Here we demonstrate a new advanced peak determination algorithm – APD – that improves precursor annotation and increases proteomic coverage. Tribrid quadrupole-Orbitrap-linear ion trap mass spectrometers are now capable of MS/MS acquisition rates exceeding 60 Hz. At these exceptional speeds, the pool for precursors available for sampling is rapidly exhausting, causing analyzer idling. This squandered instrument capacity could be utilized if more precursors could be detected in the MS¹ scans. Here we describe an algorithm (APD) that annotates more precursors in the MS¹ compared to the current standard peak detection algorithm, resulting in a near complete utilization of MS/MS capacity, increased total MS/MS scans per analysis, and improved proteomic coverage and quantitation.

The APD algorithm has several important advantages over the standard algorithm. First, the standard algorithm can only detect a single isotope distribution in a given m/z window. As a consequence, many peaks remain unannotated, especially in cases where multiple independent isotope distributions overlap within a window, even if their charge state could be determined. By contrast, APD iterates multiple times over an m/z window to find as many distinct features as possible. Second, APD performs a charge state deconvolution after the independent cluster analysis to identify multiply charged analytes across the entire spectrum, significantly improving the charge state determination of individual isotope distributions. This particularly applies to higher charge states, which often elude an unambiguous identification due to the narrow peak distances, resulting in ambiguous charge assignment scores.

Given that the additional precursors identified by APD tend to originate from regions of overlapping m/z peaks, these additional MS/MS spectra tend to have reduced success rates. That is, the new MS/MS scans generally stem from more difficult to sequence precursors. We demonstrate, however, considerably improved proteomic sampling depth because many more precursors are annotated with the APD algorithm and it is better to sample the difficult to sequence precursors than to pass them by without collecting an MS/MS spectrum. That said, there are certain situations where such improvements may be detrimental. For methods with slower MS/MS acquisition rates, the standard algorithm could identify enough precursor to keep the instrument busy. As noted above, enabling APD in these cases could have the detrimental effect of trading lower intensity but higher purity precursor for precursors of higher intensity but lower purity. As detailed in the results, this precursor exchange seems to negatively impact FTMS² analysis of phosphopeptides. We expect similar complications during MS quantification via isobaric tagging.²⁶ Here we anticipate the APD algorithm would produce similar improvements to peptide identification; however, the additional gains would mostly result from precursors with higher impurities causing potential degradation of quantitative figures of merit. We envision future APD implementations could additionally weight the purity of precursors. Additionally, improvements to chromatographic performance could increase detectable precursor by reduced co-elution of species with similar m/z and increasing overall sensitivity.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors gratefully acknowledge support from Thermo Fisher Scientific and NIH Grant R35 GM118110 awarded to J.J.C. NMR was funded through an NIH Predoctoral to Postdoctoral Transition Award (Grant F99 CA212454). The authors also thank Eugen Damoc, Mike Senko, and Philip Remes for helpful discussions.

References

1. Washburn MP, Wolters D, Yates JR III. Nature biotechnology. 2001; 19:242.
2. Peng J, Elias JE, Thoreen CC, Licklider LJ, Gygi SP. Journal of Proteome Research. 2003; 2:43–50. [PubMed: 12643542]
3. Caprioli RM, Farmer TB, Gile J. Analytical Chemistry. 1997; 69:4751–4760. [PubMed: 9406525]
4. Shevchenko A, Jensen ON, Podtelejnikov AV, Sagliocco F, Wilm M, Vorm O, Mortensen P, Boucherie H, Mann M. Proc. Natl. Acad. Sci. U. S. A. 1996; 93:14440–14445. [PubMed: 8962070]
5. Wollnik H, Przewlaka M. Int. J. Mass Spectrom. Ion Process. 1990; 96:267–274.
6. Syka JEP, Marto JA, Bai DL, Horning S, Senko MW, Schwartz JC, Ueberheide B, Garcia B, Busby S, Muratore T, Shabanowitz J, Hunt DF. Journal of Proteome Research. 2004; 3:621–626. [PubMed: 15253445]
7. Schwartz JC, Senko MW, Syka JEP. J. Am. Soc. Mass Spectrom. 2002; 13:659–669. [PubMed: 12056566]
8. Makarov A, Denisov E, Kholomeev A, Balschun W, Lange O, Strupat K, Horning S. Analytical Chemistry. 2006; 78:2113–2120. [PubMed: 16579588]
9. McAlister GC, Phanstiel D, Good DM, Berggren WT, Coon JJ. Analytical Chemistry. 2007; 79:3525–3534. [PubMed: 17441688]

10. Michalski A, Damoc E, Lange O, Denisov E, Nolting D, Müller M, Viner R, Schwartz J, Remes P, Belford M, Dunyach J-J, Cox J, Horning S, Mann M, Makarov A. *Molecular & Cellular Proteomics*. 2012; 11
11. Senko MW, Remes PM, Canterbury JD, Mathur R, Song Q, Eliuk SM, Mullen C, Earley L, Hardman M, Blethrow JD. *Analytical Chemistry*. 2013; 85:11710–11714. [PubMed: 24251866]
12. Hebert AS, Richards AL, Bailey DJ, Ulbrich A, Coughlin EE, Westphall MS, Coon JJ. *Molecular & Cellular Proteomics*. 2014; 13:339–347. [PubMed: 24143002]
13. Richards AL, Hebert AS, Ulbrich A, Bailey DJ, Coughlin EE, Westphall MS, Coon JJ. *Nat. Protocols*. 2015; 10:701–714. [PubMed: 25855955]
14. Wilhelm M, Schlegl J, Hahne H, Gholami AM, Lieberenz M, Savitski MM, Ziegler E, Butzmann L, Gessulat S, Marx H. *Nature*. 2014; 509:582. [PubMed: 24870543]
15. Kim M-S, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaerkady R, Madugundu AK, Kelkar DS, Isserlin R, Jain S. *Nature*. 2014; 509:575. [PubMed: 24870542]
16. Aebersold R, Mann M. *Nature*. 2016; 537:347–355. [PubMed: 27629641]
17. Richards AL, Merrill AE, Coon JJ. *Current Opinion in Chemical Biology*. 2015; 24:11–17. [PubMed: 25461719]
18. Riley NM, Hebert AS, Coon JJ. *Cell systems*. 2016; 2:142–143. [PubMed: 27135360]
19. Nagaraj N, Alexander Kulak N, Cox J, Neuhauser N, Mayr K, Hoerning O, Vorm O, Mann M. *Molecular & Cellular Proteomics*. 2012; 11
20. Kelstrup CD, Jersie-Christensen RR, Batth TS, Arrey TN, Kuehn A, Kellmann M, Olsen JV. *Journal of Proteome Research*. 2014; 13:6187–6195. [PubMed: 25349961]
21. Michalski A, Cox J, Mann M. *Journal of Proteome Research*. 2011; 10:1785–1793. [PubMed: 21309581]
22. Shishkova E, Hebert AS, Coon JJ. *Cell systems*. 2016; 3:321–324. [PubMed: 27788355]
23. Blue LE, Franklin EG, Godinho JM, Grinias JP, Grinias KM, Lunn DB, Moore SM. *Journal of Chromatography A*. 2017; 1523:17–39. [PubMed: 28599863]
24. Horn DM, Zubarev RA, McLafferty FW. *J. Am. Soc. Mass Spectrom*. 2000; 11:320–332. [PubMed: 10757168]
25. Senko MW, Beu SC, McLafferty FW. *J. Am. Soc. Mass Spectrom*. 1995; 6:229–233. [PubMed: 24214167]
26. Wenger CD, Lee MV, Hebert AS, McAlister GC, Phanstiel DH, Westphall MS, Coon JJ. *Nat Meth*. 2011; 8:933–935.

Figure 1.

Basic workflow of the advanced peak determination algorithm. The three algorithm steps build upon each other to efficiently analyze charge states in local isotope clusters (step II) and to correlate isotope distributions originating from multiply charged analytes (step III). Lists of isotope distributions and deconvolved masses forming charge envelopes in the spectrum are obtained as the result of the analysis.

Figure 2. Annotation of MS¹ spectra of three overlapping isotope clusters with and without APD
a) With standard peak determination (top), only the green cluster ($z = 3$) and a subset of the blue cluster are correctly annotated. The missed or incorrect annotations are highlighted in grey and the charge states can be easily calculated with the noted m/z values. The APD algorithm (bottom) correctly annotates all 3 clusters observed in the example spectra. **b)** The same clusters are observed 1–1.4 seconds later in the gradient now with the blue cluster being the most abundant. Without using the APD algorithm, only the most abundant cluster is annotated while APD again correctly annotates all three species.

Figure 3. Maximizing precursor sampling capacity with the APD algorithm

When acquiring data with an MS/MS scan rate of approximately 35 Hz, the standard peak determination algorithm utilizes only 72% of the available sampling capacity (a) because the instrument runs out of annotated precursors to select. APD (b), on the other hand, nearly maximizes sampling capacity by providing more annotated precursor ions to select for MS/MS fragmentation.

Figure 4. The APD algorithm enables sampling and identification of more peptide features
a) Standard peak determination (top) allows selection of only a fraction of the available peptide features for MS/MS analysis. The APD algorithm (bottom) significantly increases the number of features sampled and identified via MS/MS, especially in the most densely populated regions of the MS^1 spectra (i.e., 400–900 m/z). **b)** The SPD algorithm inadvertently limits the MS/MS sampling of impure peptides. Peptide selection is less affected by purity when using APD. The success rate (identified/sampled) for each bin from the APD data is plotted in red **c)** Because of the crude filtering function in SPD, when spectral acquisition speeds are low enough that precursors are never exhausted, precursors selected for MS/MS will have greater dynamic range. The combination of finding more precursors, with APD, and faster spectral acquisition leads to more comprehensive sampling.

Figure 5. The APD algorithm improves peptide and protein identifications over the standard algorithm

a) Improved precursor annotation with APD results in more MS/MS scans, as more precursor ions meet the charge state criterion for selection. With APD, higher MS^1 resolutions correlate with increased MS/MS events. **b)** The APD algorithm shows significant benefits in peptide identification at 120K and 240K MS^1 resolutions because the instrument can perform ion trap MS/MS in parallel while taking advantage of the improved spectral MS^1 quality from the longer transients. **c)** Even with longer gradients (i.e., higher peak capacity), the APD algorithm enables more unique peptide identifications. **d)** The benefits of APD translate to protein identifications, as well, including the analysis of pre-fractionated samples. Note, the y-axes do not start at zero and error bars represent minimum and maximum values for two replicates.

Figure 6. Performance of the APD algorithm for multiple proteases and phosphopeptides
a) More unique peptides from mouse brain tissue are identified using the APD algorithm for each of the three alternative proteases investigated. Error bars represent minimum and maximum values of two replicates. Run times were 60 minutes for each analysis. **b)** When acquiring MS/MS scans in the ion trap (*i.e.*, when the instrument can parallelize MS¹ and MS/MS acquisitions), the APD algorithm enables identification of ~15% more unique phosphopeptide sequences. When acquiring MS/MS scans in the Orbitrap, however, (*i.e.*, when the instrument is operating at lower scan speeds) the standard method outperforms APD. Run times were 3 hrs for each analysis.

Figure 7. APD gains depend upon spectral acquisition rate

a) Average ($n=2$) FT MS/MS spectra collected and % improvement ($((APD-SPD)/SPD)$) as a function of FT MS/MS resolving power using APD or SPD. **b)** Average ($n=2$) unique peptides identified (1% FDR) and % improvement ($((APD-SPD)/SPD)$) as a function of FT MS/MS resolving power using APD or SPD. APD produces the greatest increases in both MS/MS scans and unique peptides at fastest acquisition rates (lowest resolving power).