Exploring the effects of alternative dynamic exclusion algorithms on peptide identification experiments

Nina L. Soltero, Graeme C. McAlister, Derek J. Bailey, Vlad Zabrouskov, Thermo Fisher Scientific, 355 River Oaks Parkway, San Jose, CA, USA, 95134

INTRODUCTION

Bottom-up experiments for peptide identification are a common part of proteomics studies. Often, the objective of these experiments is to identify as many peptides as possible from a particular sample of interest. For samples of high concentration, complexity, and dynamic range, effective instrument control is essential. Utilization of dynamic exclusion (DE) lists is a key aspect of efficient data acquisition in data-dependent (DDA) workflows when analyzing samples of such complexity. These method filters determine on-line which peaks to exclude from subsequent MS2 interrogation. The filter settings - such as the exclusion duration, repeat count, and the choice of algorithm - can have a significant impact on the results of the experiment. Here, we explore the effects of these parameters on the results of a bottom-up peptide identification experiment.

BACKGROUND

Dynamic exclusion lists were developed to prevent repeat MS^2 acquisitions of high abundance precursors. With the speed and sensitivity of current mass spectrometers, dynamic exclusion algorithms should be precise and selective enough to sample the maximum amount of different species in every scan cycle. However we have empirically found that repeat acquisitions, especially for lower abundance precursors, can be advantageous. In some cases a less restrictive dynamic exclusion list can produce a greater number of unique peptide identifications. Analyses have shown that oftentimes we are not maximizing the potential MS^2 acquisition rate and are running out of precursors to interrogate (Figure 1). The way in which we utilize the extra "idle time" on more precursors may be critical to the outcome of standard bottom-up peptide identification experiments. Our study was based upon observations of two different dynamic exclusion algorithms (Figures 2 through 4).

Figure 1. MS2 acquisition rate

Plot of observed MS^2 acquisition rate (blue) compared to maximum MS^2 acquisition rate (red) for OTMS-CID-ITMS^2 experiments on 1 ug HeLa.

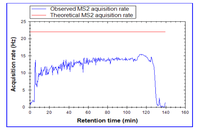
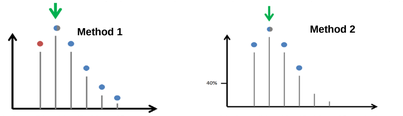


Figure 2. Two different dynamic exclusion algorithms for determining peaks assigned to dynamic exclusion (DE) list

MS/MS candidate indicated by green arrow. Blue represent peaks effectively excluded for entirety of exclusion duration period with each DE method. Gray represents the apex of the isotopic envelope. Red represents mono-isotopic peak.



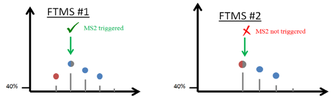
- MS/MS candidate identified for MS^2 analysis
- Algorithm determines mono-isotopic peak of MS/MS candidate's cluster
- Mono-isotopic peak placed on DE list
- Next MS/MS candidate identified for MS^2 analysis
- Algorithm determines mono-isotopic peak of MS/MS candidate's cluster
- If the mono-isotopic peak matches any peaks previously placed on DE list, MS^2 event will not be triggered

- MS/MS candidate identified for MS^2 analysis
- Algorithm determines rest of peaks that make up MS/MS candidate's isotopic envelope and determines apex of MS/MS candidate's cluster
- Apex of MS/MS candidate's cluster is placed on DE list, along with any isotopes in the cluster that have an intensity > 40% of the apex
- Next MS/MS candidate identified for MS^2 analysis
- Algorithm determines apex of MS/MS candidate's cluster
- If the apex peak matches any peaks previously placed on DE list, MS^2 event will not be triggered

Figure 3. Analysis of low-level features using Method 1

An artifact of method 1 is that the mono-isotopic peak may be mis-assigned, i.e. in cases of low abundance peaks (Figure 3). In this example, in FTMS #1, the true mono-isotopic peak is intense enough to be detected in the spectrum, enabling correct mono-isotopic peak assignment of the MS/MS candidate and subsequent ddMS^2, with the mono-isotopic peak placed on the DE list. However, in the next scan (FTMS #2), the true mono-isotopic peak is not detected in the spectrum and the monoisotopic peak is incorrectly assigned. This peak is triggered on again for MS^2 with the incorrect mono-isotopic peak placed on the DE list. Method 2 was developed in an effort to alleviate the issue of multiple triggers.

Figure 4. Analysis of low-level features using Method 2



In the case of low level clusters that may have mis-assigned mono-isotopic peaks, the algorithm is unaffected since it only compares the m/z of the MS/MS candidate's cluster apex and its associated peaks to the dynamic exclusion list. Method 2 is more stringent than Method 1 because it does not allow for multiple triggers.

MATERIALS AND METHODS

In this study, we analyzed 1 ug of a commercially available HeLa protein digest from Pierce (Rockford, IL) with the Thermo Scientific™ EASY-nLC 1000 ultra-high pressure LC and a Thermo Scientific™ Orbitrap Fusion™ Lumos™ Tribrid™ Mass Spectrometer (San Jose, CA). We first examined the effects of the two different dynamic exclusion methods on standard FTMS-CID-ITMS^2 bottom-up data-dependent MS^2 experiments on 1 ug HeLa digest over a 2-hour gradient as previously described [1], except with a 25-cm column. We ran these experiments in duplicate with 15 and 60 second exclusion durations (Experiment 1). We repeated these experiments with a shallower LC gradient to mimic poorer chromatography; i.e. broader peaks (Experiment 2). We then performed FTMS-HCD-OTMS^2 data-dependent experiments with the optimized LC gradient (Experiment 3). The resulting LC-MS/MS data were searched using the Thermo Scientific™ Proteome Discoverer™ software. The spectra were searched against the Uniprot Human database and the peptide spectral matches were filtered to a 1% false discovery rate. The data was subsequently analyzed with in-house tools [2] to further understand how the dynamic exclusion filter affected peptide and protein identifications, success rates, and other parameters of interest for peptide/protein identification experiments.

RESULTS

Experiment 1. For runs using standard dynamic exclusion conditions (DE = 60 seconds), we found that method 1 resulted in a higher number of identifications (Figure 5). This was accompanied by a proportional increase in number of MS^2 triggers and number of peptide spectral matches (PSMs). For the experiments using DE = 15 seconds, the number of peptide identifications, PSMs and MS^2 triggers between the two dynamic exclusion methods were similar (Figure 6).

The results are expected because lowering DE times results in acquiring more MS^2 over an entire run. We have empirically found that peptide identifications increase with shorter DE times (i.e. 15 seconds) because it better matches our LC peak widths with our current gradient and enables more repeat analyses of the same precursor (Figure 7). The difference between DE algorithms with different levels of stringency depends on the DE settings. Low DE times essentially negates the effect of dynamic exclusion, which is intended to prevent the same peak from being interrogated within the exclusion duration and enables the instrument to sample other precursors. The similar results between both algorithms when utilizing DE = 15 seconds in peptide identifications, PSMs and MS^{2}s are as expected. On the other hand, longer exclusion duration times results in a larger divergence in the performance between the two DE algorithms. This is because the specificity of the DE algorithm is affected by things like correct mono-isotopic peak assignment, but a longer exclusion duration yields fewer chances for repeat analyses. Accordingly, when the DE = 60 seconds, method 2 results fewer peptide identifications than method 1, because of the stringency of its peak exclusion algorithm. The decrease in peptide identifications (around 3%) is accompanied by a proportional decrease in MS^2 and PSMs.

Figure 5. Peptide identifications, PSMs and MS^2 totals from OTMS-CID-ITMS^2 experiments with DE = 60 s

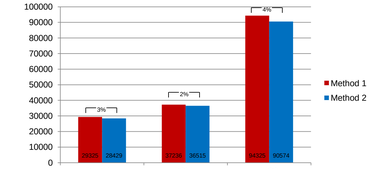


Figure 6. Peptide identifications, PSMs and MS^2 totals from OTMS-CID-ITMS^2 experiments with DE = 15 s

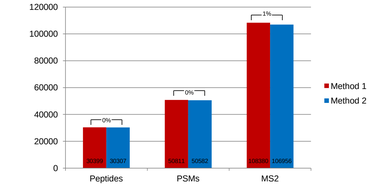
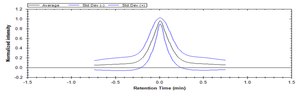


Figure 7. Average normalized intensity vs. retention time of all LC-MS features detected in a single run



For the subsequent analysis (Figures 8 and 9), we restricted the dataset to the MS^2 scans that were associated with an LC feature (based on an analysis of XIC for that particular precursor and its isotopes). In our analysis of trigger order (the number of times a particular precursor is triggered on for MS^2 across the entire LC-MS run), we found that for both DE settings, there were more higher-order repeat triggers using method 1, while there were more precursors interrogated just once with method 2.

Figure 8. Number of scans for each trigger order in a OTMS-CID-ITMS^2 peptide identification experiment with dynamic exclusion duration set to 60 seconds.

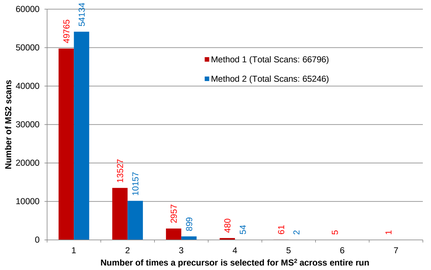
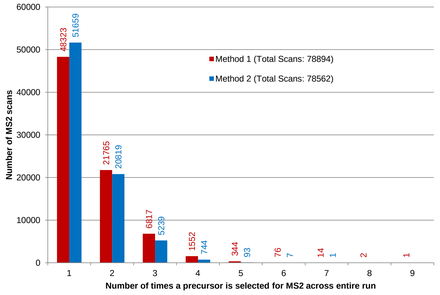
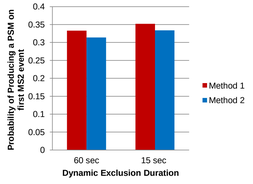


Figure 9. Number of scans for each trigger order in a OTMS-CID-ITMS^2 peptide identification experiment with dynamic exclusion duration set to 15 seconds.



Looking closer at the features interrogated a single time, we found that the probability of producing a PSM is higher with method 1 than method 2 (Figure 10). This is expected since method 2 is more selective and will force the instrument to start analyzing very low-level features, which may result in subpar MS^2 quality, due to lack of adequate signal or high levels of interference, and thus no peptide-spectral match.

Figure 10. Probability of producing a PSM on the first trigger in OTMS-CID-ITMS^2 experiments



For this restricted data set, we calculated the time relative to the peak apex (t = 0) of the first MS^2 event (Table 1) and observed slight differences between the two algorithms. For both DE settings, method 2 triggered slightly earlier. We also calculated the average precursor intensity of the first MS^2 event (Table 2) and measured higher average precursor intensities using method 1 compared with method 2 for both DE settings, which is expected due to the higher selectivity of method 2, which induces the instrument to go after more lower-level precursors.

Table 1. Average MS^2 retention time of 1st MS^2 event in OTMS-CID-ITMS^2 peptide identification experiments

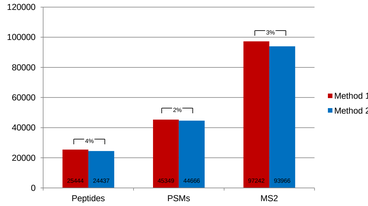
	Average MS^2 Retention Time	
	Method 1	Method 2
DE = 60 s	-9.96 seconds	-12.00 seconds
DE = 15 s	-9.66 seconds	-10.38 seconds

Table 2. Average precursor intensity of 1st MS^2 event in OTMS-CID-ITMS^2 peptide identification experiments

	Average Precursor Intensity	
	Method 1	Method 2
DE = 60 s	1202264	959400
DE = 15 s	1479108	1218989

Experiment 2. We next assessed the performance of the two algorithms on 1 ug HeLa experiments with the same instrument method (OTMS-CID-ITMS^2) parameters and DE = 15 seconds, but with a shallower LC gradient to mimic poorer chromatography (i.e. broad peaks). As expected, the overall peptide identifications, PSMs and MS^{2}s are lower than for the corresponding data shown in Experiment 1 for the optimized gradient. In this case, we observed differences in overall peptide identifications between the two DE methods (Figure 11), while in Experiment 1 (optimized LC gradient and DE = 15 s), both methods yielded similar numbers of peptide identifications, PSMs and MS^{2}s.

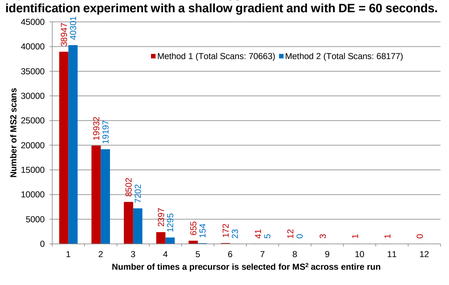
Figure 11. Peptide identifications, PSMs and MS^2 totals from OTMS-CID-ITMS^2 experiments with DE = 15 s, using shallow gradient



We repeated the analysis of MS^2 scans correlated with LC peaks (Figure 12). Expectedly, we saw more higher-order repeat triggers in this experiment than in experiment 1 because, with a shallower gradient, peaks elute over a longer period of time and are available longer for repeat sampling. Similar to experiment 1, method 1 resulted in a higher numbers of multiple MS^2 scan events for the same feature.

We also observed differences in the relative performance between the 2 DE algorithms in terms of single triggers (features only analyzed once). In experiment 1 with DE = 15 seconds, there were almost 7% more single triggers with method 2 than method 1, while here the difference is < 3.5%. In Experiment 1, the additional interrogation of many new features helped bring the performance of method 2 to the level of method 1 by analyzing many more unique precursors. However, in this experiment with poorer chromatography, there are not a significant amount of new precursors being selected for MS^2. The less stringent method 1 performs better due to having more repeat analyses of previously selected precursors, which enhances the probability of a PSM and thus increases peptide identifications. This is further confirmed by the fact that the probability of producing a PSM on a single trigger for both methods does not change with the shallow gradient, and is higher with method 1 (Figure 13).

Figure 12. Number of scans for each trigger order in a OTMS-CID-ITMS^2 peptide identification experiment with a shallow gradient and with DE = 60 seconds.



Experiment 3. We repeated the analysis of the different dynamic exclusion methods on 1 ug HeLa and with FTMS-HCD-FTMS^2 methods and the 2-hour optimized LC gradient to confirm these findings were not limited to the type of instrument method we used for bottom-up experiments. We found similar results to that of the FTMS-CID-ITMS^2 methods; namely that utilization of method 1 triggers more MS^2 and thus yields higher peptide identifications than using method 2 (Figures 14 and 15). The difference is more pronounced in the experiments where the dynamic exclusion duration was set to 60 seconds, where dynamic exclusion plays a larger role in modulating the number of scans taken and thus the unique peptides identified.

Figure 13. Probability of producing a PSM on the first trigger in OTMS-CID-ITMS^2 experiments with DE = 15 s

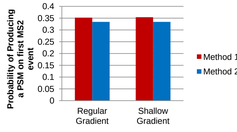


Figure 14. Peptide identifications, PSMs and MS^2 totals from OTMS-HCD-OTMS^2 experiments with DE = 60 s

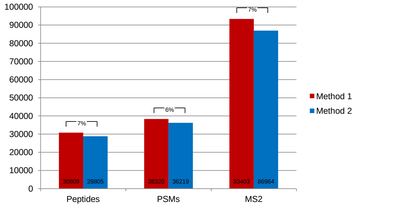
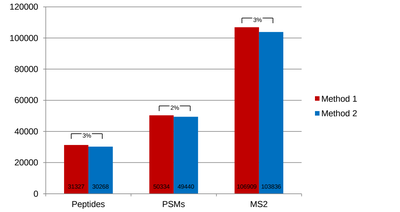


Figure 15. Peptide identifications, PSMs and MS^2 totals from OTMS-HCD-OTMS^2 experiments with DE = 15 s



CONCLUSIONS

Overall, the data demonstrates that while the intensity-based method is more specific, results in fewer acquisitions and interrogates lower-level precursors, it results in a lower number of PSMs and unique peptide identifications in most bottom-up peptide identification experiments. The data show that small differences in the approach to peak exclusion in dynamic exclusion algorithms can have a small but consistent effect on peptide identification results. We also show that using method 1 can improve results in cases of poorer chromatography. When peaks get broader and elute over a longer period of time, features that may have been present on only a couple of FTMS scans in experiments with optimized chromatography may be present in many more FTMS scans and have a higher chance of mono-isotopic peak mis-assignment in one of these scans. This will then increase the number of features interrogated for MS^2 due to repeat analyses with method 1, and subsequently yield higher numbers of PSMs and peptide identifications.

REFERENCES

1. Lopez-Ferrer D, Blank M, et al. **Pushing the Limits of Bottom-Up Proteomics with State-Of-The-Art Capillary UHPLC and Orbitrap Mass Spectrometry for Reproducible Quantitation of Proteomes**. Thermo Fisher Application Note 639
2. Senko M, McAlister GC, Bailey D. **Xcalibur Workbench: A Lua Based Data Browser.** ASMS Poster 2016 WP404

TRADEMARKS/LICENSING