# Integrating Filtered Peak Lists into Protein Prospector's Batch-Tag

Peter R. Baker[1], Clinton Yu[2] and Robert J. Chalkley[1]

[1]Mass Spectrometry Facility, Dept. of Pharmaceutical Chemistry, University of California, San Francisco, USA

[2]Department of Physiology & Biophysics, University of California, Irvine, CA 92697

## Introduction

Protein Prospector has a peak list filtering program called MS-Filter which can select a subset of MSMS peptide spectra based on criteria such as the precursor charge state and fragment m/z values or differences present. One or more peak lists are uploaded and a generally smaller filtered set of peak lists can be downloaded. Alternatively the spectra matching the criteria can be removed from the initial peak lists. Typical uses of this would be to extract spectra with glycan-specific oxonium ions or spectra containing crosslinks. Filtered peak lists can improve the speed and specificity of searches. MS-Filter has now been integrated directly into Batch-Tag searching.

## Methods

Filtered peak lists from MS-Filter can now be part of a Batch-Tag project and are available as a parameter on the Batch-Tag form. Rather than storing the filtered peak lists separately the filtering is stored as a file containing a set of indices to the retained or excluded spectra. This saves a lot of disk space and means the peak lists don't have to be uploaded or downloaded. Also, sets of filtered spectra can be combined using union, difference and intersection set operators. In addition Search Compare now has an option for creating a filtered peak list based on matched or unmatched spectra.

Fig. 1 shows the MS-Filter form and the options available. An optional report is available with details of the matched spectra (Fig. 2). MS-Filter can be applied in simple cases such as eliminating spectra with high precursor charges or selecting spectra with certain glycosylation oxonium ions.(Fig. 3).

There are also more complex options such as looking for crosslinker-specific ions (Fig. 4 and Fig.5) or glycosylation Y ions which match a list of glycopeptides and a set of glycosylation modifications[1].

## Conclusions

• The Protein Prospector MS-Filter program for producing filtered peak lists has been integrated into Batch-Tag searching. The filtered peak lists are represented by a list of index numbers rather than the peak lists themselves. This saves a lot of disk space and means the lists can be combined using set operators.

• A sophisticated use of MS-Filter to detect glycosylation has been previously described[1]. The poster "Increasing Glycopeptide Identifications in Large-Scale Studies through Targeted Y Ion Detection using MS-Filter" outlines some recent developments. This has been the most successful application of MS-Filter to date.

• MS-Filter can attempt to detect spectra of cleavable crosslinked peptides based on fragment ion pair delta masses. This was found to be helpful in certain cases but needs further development.

• Search Compare can also produce filtered peak lists of matched or unmatched spectra allowing datasets with different types of spectra (eg linear and crosslinked peptides) to be searched in stages.

• A report showing details about the matched spectra is available if required.

## References

1. Chalkley R. J., Medzihradszky K. F., Pap A., Darula Z., Baker P. R. (2020) The effectiveness of filtering glycopeptide peak list files for Y ions. Molecular Omics Vol. 16, No 2. Pp. 147-155

## Acknowledgements

## MS-Filter Options



**Fig. 1 MS-Filter options**

Fig. 1 shows the current MS-filter options, The available filters include:

• Precursor M+H and charge .
• Retention time.
• One or more fragment m/z values with a minimum number of matches
• Up to 2 m/z difference with a minimum number of occurrences.
• A loss composition
• Glycosylation filters[1].
• Options related to cleavable crosslinkers.

## Filtering Cleavable Crosslinked Spectra

• Cleavable crosslinkers such as DSSO produce pairs of peaks for each peptide backbone fragmentation; an alkene (A) and thiol (T)
•MS-Filter can search MSMS spectra to find spectra that contain pairs of peaks separated by the mass difference between A and T fragments:
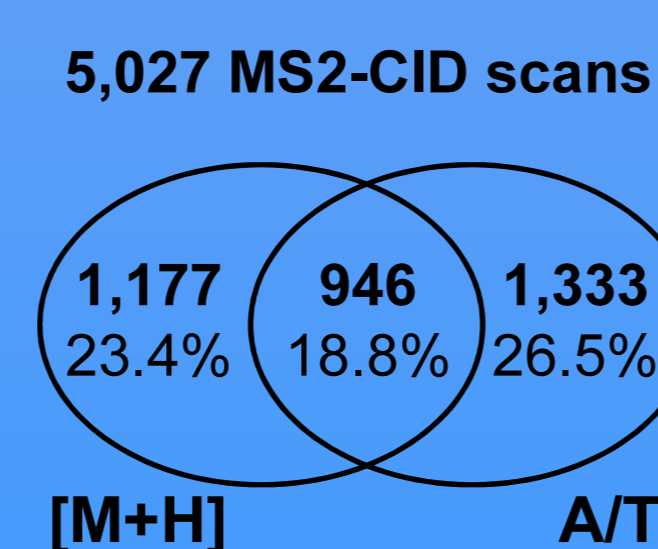


**5,027 MS2-CID scans**

1,177 23.4% | 946 18.8% | 1,333 26.5%

[M+H] — A/T

**4,990 MS2-HCD scans**

828 16.6 | 779 15.6% | 439 8.8%

[M+H] — A/T

**Fig. 4 Scans passing MS-Filter:** 2,046 (68.7%) CSMs retained: 65/65 (100%) 0% IDs lost. Spectra reduced by 31%.
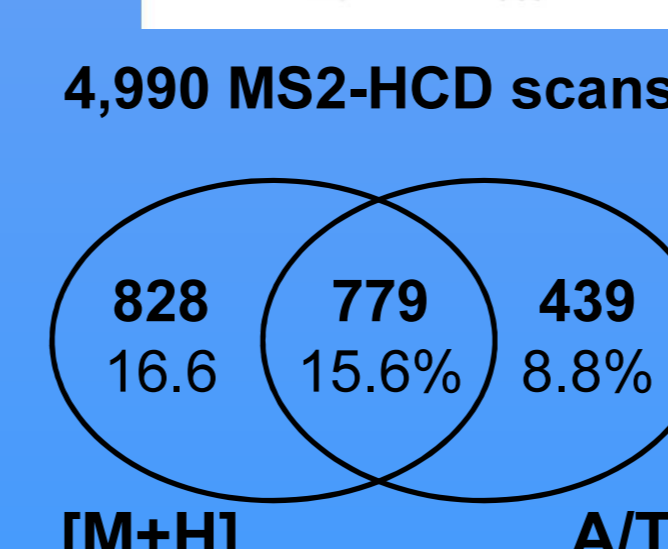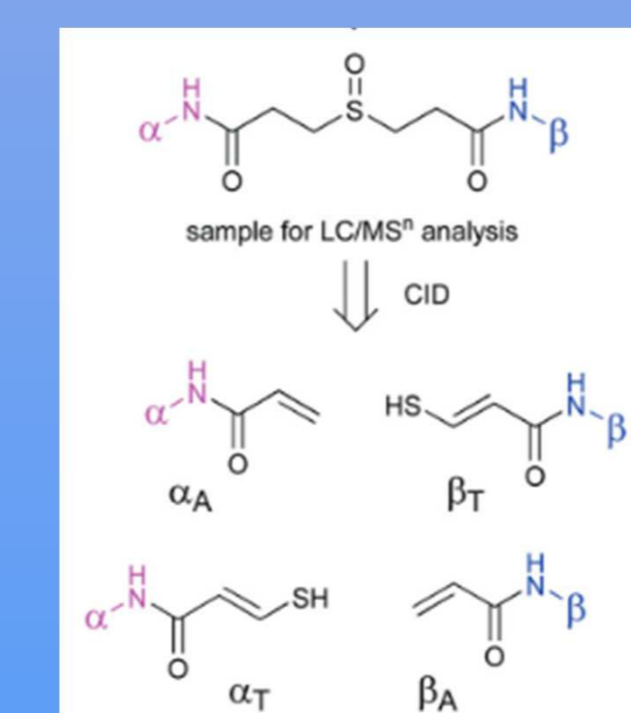
**Fig. 5 Scans passing MS-Filter:** 2,046 (41.0%) CSMs retained: 64/78 (82%) 18% IDs lost. Spectra reduced by 59%.

• MS-Filter is effective for filtering CID-MS scans, but not ideal for filtering HCD-MS2 scans in its current iteration.

## Filtered Spectrum Report

An optional report is available giving details of the spectra that have been retained or excluded. The report can be sorted by up to 6 levels. An example is given below.

| | 1 | 2 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Index | m/z | z | M+H | Intensity | Fraction | RT | MSMS Info | Num Pks | Max Intensity | Num 2 | Num 3 | Num 4 | Max Z |
| 1 | 495.8333 | 3 | 1485.4853 | 48452 | Z20200203-10 | 0.100 | 17 | 24 | 11500 | 0 | 0 | 0 | 1 |
| 2 | 419.3159 | 2 | 837.6244 | 31339 | Z20200203-10 | 0.167 | 26 | 16 | 3323 | 0 | 0 | 0 | 1 |
| 3 | 716.6373 | 3 | 2147.8974 | 23677 | Z20200203-10 | 4.583 | 608 | 33 | 2956 | 9 | 0 | 0 | 2 |
| 4 | 824.6731 | 3 | 2472.0047 | 16667 | Z20200203-10 | 4.700 | 623 | 11 | 1120 | 1 | 0 | 0 | 2 |
| 5 | 770.6549 | 3 | 2309.9502 | 24460 | Z20200203-10 | 4.700 | 624 | 15 | 1901 | 4 | 0 | 0 | 2 |
| 6 | 932.7074 | 3 | 2796.1076 | 16233 | Z20200203-10 | 4.717 | 626 | 25 | 1289 | 3 | 0 | 0 | 2 |
| 7 | 1074.4524 | 2 | 2147.8975 | 18040 | Z20200203-10 | 4.717 | 628 | 16 | 1550 | 3 | 0 | 0 | 2 |
| 8 | 729.2858 | 3 | 2185.8428 | 11996 | Z20200203-10 | 4.733 | 630 | 16 | 1932 | 7 | 0 | 0 | 2 |
| 9 | 662.6188 | 3 | 1985.8420 | 22296 | Z20200203-10 | 4.750 | 632 | 34 | 2122 | 8 | 0 | 0 | 2 |
| 10 | 878.6905 | 3 | 2634.0569 | 19816 | Z20200203-10 | 4.833 | 643 | 22 | 1171 | 8 | 0 | 0 | 2 |
| 11 | 993.4264 | 2 | 1985.8455 | 17495 | Z20200203-10 | 4.883 | 649 | 21 | 2035 | 4 | 0 | 0 | 2 |
| 12 | 735.2762 | 3 | 2203.8140 | 12116 | Z20200203-10 | 4.917 | 654 | 19 | 962.3 | 5 | 0 | 0 | 2 |
| 13 | 716.6379 | 3 | 2147.8991 | 1.103e+6 | Z20200203-10 | 5.083 | 677 | 40 | 68187 | 14 | 0 | 0 | 2 |

**Fig. 2 Report showing details of the filtered spectra.**

## Filtering for Phospho Glycan Oxonium Ions

The oxonium ions for mannose-6-phosphate-containing N-glycopeptdes are at m/z 243.0264 and 405.0793. Filtering 75223 glycopeptide spectra for both these peaks reduces the data set to 309 spectra. This resulted in identifying 19 glycopeptides with phospho-containing modifications. No other glycopeptides or non-glycosylated peptides were found in the search. An example assignment is shown below.
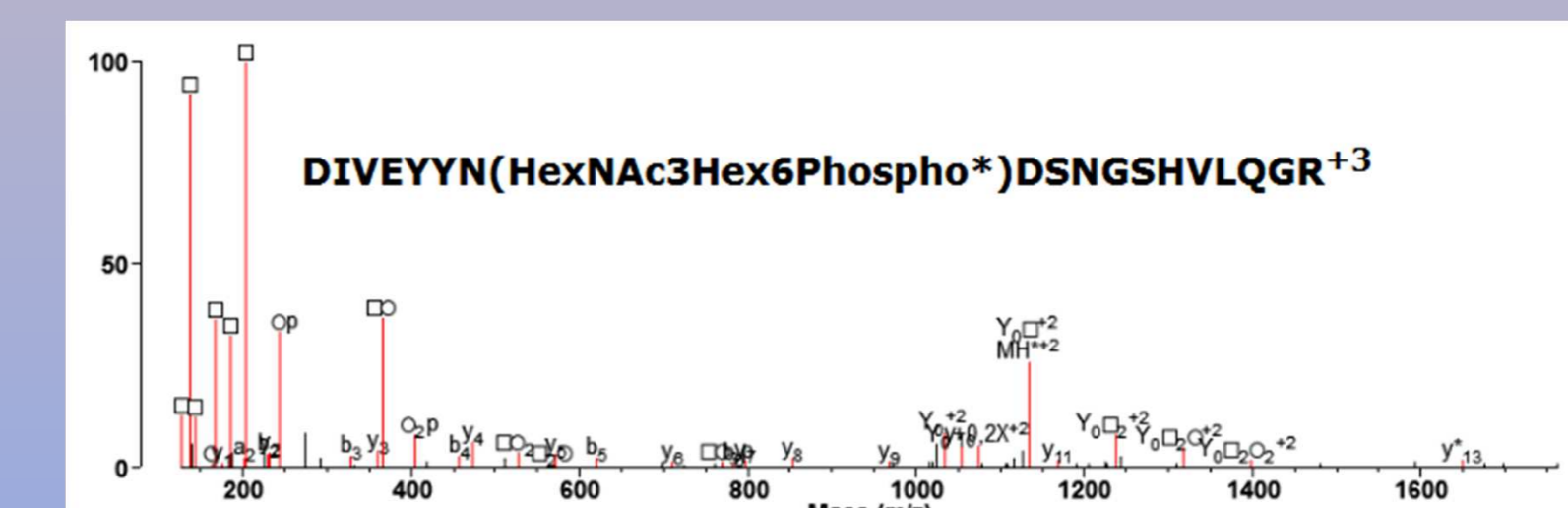


**Fig. 3 An example phospho-containing N-glycopeptide HCD spectrum. The asterisk indicates a labile modification.**

## 3 Stage Crosslinking Search

When searching for crosslinked peptides it is desirable to use different thresholds for calculating the FDR of linear peptides, intramolecular crosslinked peptides and intermolecular crosslinked peptides. One possible way around this is a 3-stage search:

• Search 1 has default and dead-end crosslink modifications and generates a list of accession numbers and matched spectra.
• Search 2 looks for intramolecular modifications. It is limited to the accession numbers found in the Search 1. The spectra matched in Search 1 are eliminated.
• Search 3 looks for intermolecular modifications. It is limited to the accession numbers identified in the Search 2. The spectra identified in Searches 1 and 2 are eliminated.

Eg. Search 1 started off with 28748 spectra. It identified1308 proteins with 5183 spectral matches (1% protein/1% peptide).
Search 2 found 398 intermolecular crosslinked proteins an 1748 crosslinked peptides. There was a 5% FDR using default Search Compare parameters except for a min score difference of 6. 398 proteins with intra. 1748 crosslinked peptides (69 Single decoy, 25 Double Decoy).
Search 3 was not very successful as there were a lot of high scoring decoy hits. At 5% FDR only 35 crosslinked peptides were detected.