

IMPROVING THE SPEED OF MULTI MODIFICATION SEARCHES IN SPECTROMINE BY ADDING AN MS2-BASED FILTER

Lynn Verbeke¹, Oliver M. Bernhardt¹, Jan Muntel¹, Lukas Reiter¹

1) Biognosys AG, Wagistrasse 21, 8952 Schlieren (Zurich), Switzerland

Lynn Verbeke, MSc
Bioinformatics Manager

lynn.verbeke@biognosys.com
www.biognosys.com



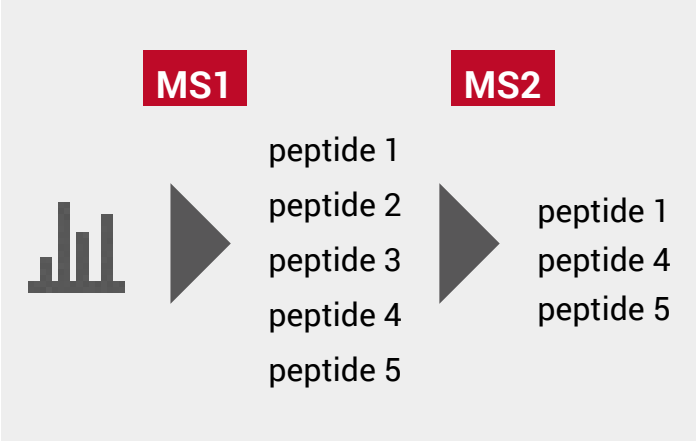
INTRODUCTION

In a typical shotgun experiment, a large amount of the spectra will not be identified. This “dark matter” remains unexplained due to peptides and post-translational modifications not in the search space^{1,2} or insufficient quality of spectra. Search engines are often not suited for searches with many modifications (memory problems, long search times) due to the combinatorial explosion of the search space. Others have described

using a fragment ion-based instead of peptide-based search space^{1,3}. In Pulsar we have now added an MS2-based filter to drastically speed up many modification searches without losing identifications. Pulsar is a versatile database search algorithm, used for library generation and directDIA in Spectronaut (Biognosys) and for isobaric labeling quantification or qualitative analysis in SpectroMine (Biognosys).

Figure 1: MS2-based filter in Pulsar

In the Pulsar pipeline, after matching a scan with all theoretical peptides within its tolerance, we have now integrated an MS2-based filter that removes all peptides with a lack in MS2 evidence (less than 3 fragments). This reduces the number of Pulsar scores to be calculated. We assumed that this would speed up searches without losing identifications.



CONCLUSIONS

- Adding the MS2-based filter in the Pulsar pipeline resulted in a very significant speed-up without losing identifications.
- This can help to identify biologically relevant modifications, but also modifications related to sample storage or preparation.

References

1. Kong, A., Leprevost, F., Avtonomov, D. et al. Nat Methods 14, 2017.
2. M. den Ridder, P. Daran-Lapujade, M. Pabst. FEMS Yeast Research 20, 2020.
3. Rinner, O., Seebacher, J., Walzthoeni, T. et al. Nat Methods 5, 2008.
4. F. Coscia, S. Doll, J. M. Bech, A. Mund, E. Lengyel, J. Lindebjerg, G. I. Madsen, J. M. A. Moreira, M. Mann. bioRxiv, 779009, 2019.

RESULTS

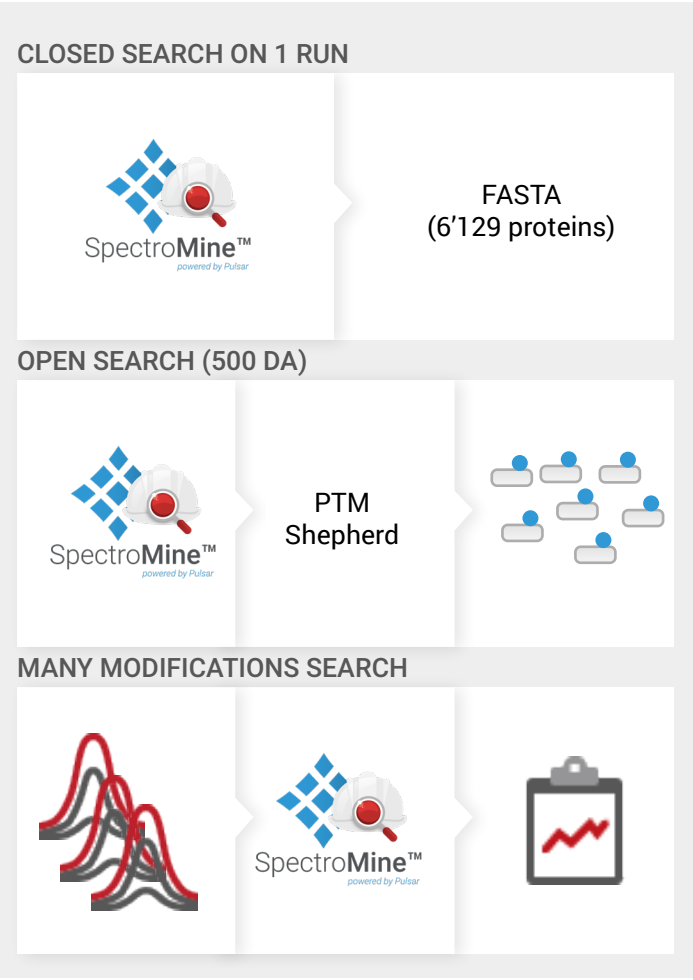


Figure 2: PTM Workflow

We applied this workflow on a formalin-fixed paraffin-embedded (FFPE) lung data set, consisting of twice 10 fractions (healthy and cancer pool), each generated from 3 samples in 3 sample preparation replicates.

1. We searched the first fraction of the lung cancer pool (default settings) to generate a smaller FASTA file.

2. We performed an open search with Pulsar (precursor mass tolerance: 500 Da). PTM-Shepherd (Nesvizskii et al, 2019) mapped the delta masses of the results to Unimod modifications, of which the 10 most common are: acetylation, amidation, ammonia loss, carbamidomethylation, carbamylation, cation:K, deamidation, formylation, methylation, oxidation.

3. Finally, we performed a closed search against the human UniProt FASTA, using the modifications as variable, setting the maximum number of modifications in 1 peptide to 4.

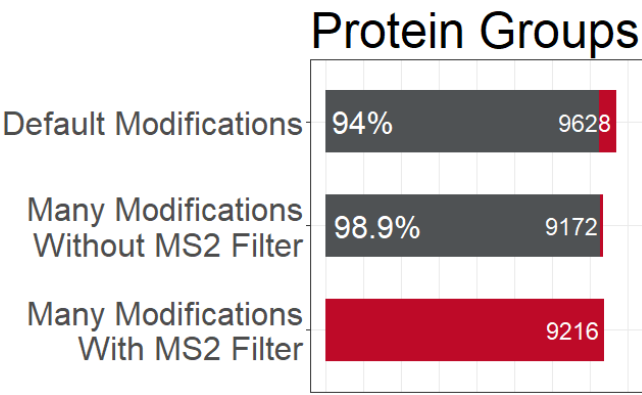
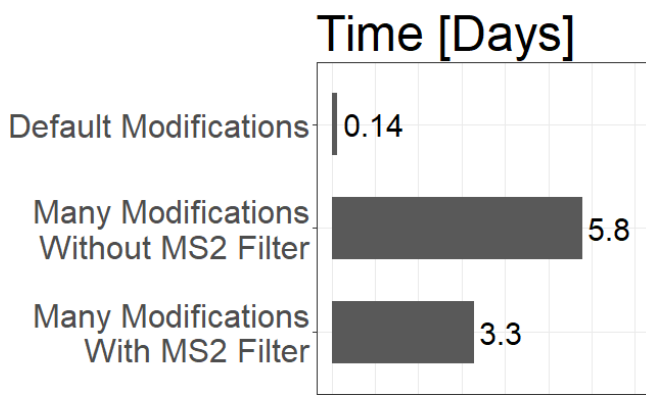
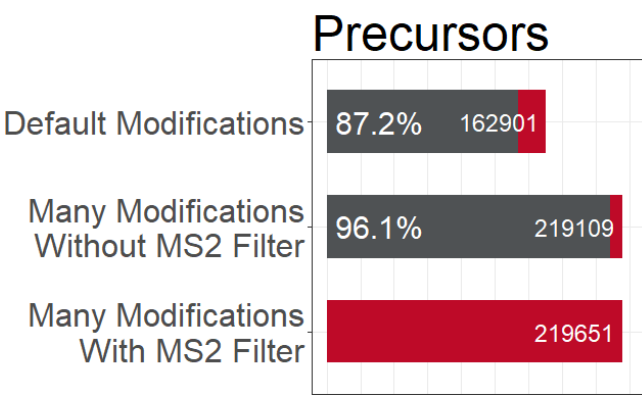


Figure 3: Results

1, 2: Identified precursors and protein groups, overlap to “Many Modifications - with MS2 Filter” in grey. The overlap is very high. Using the MS2 filter does not affect identifications, and the additional modifications help identifying previously unidentified scans.

3: Execution time in days: using the MS2-based filter speeds up the search with almost 50%.

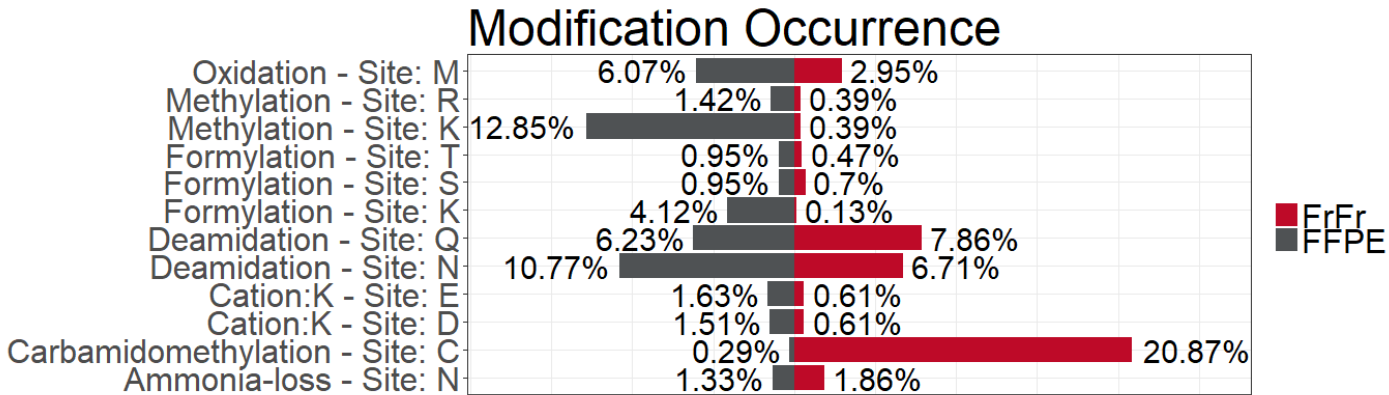


Figure 4: Modification Occurrence

We also searched fresh frozen (FrFr) lung tissue with the same settings and compared modification occurrences normalized by the number of modified peptides. We only plotted modifications with an occurrence > 0.7% in at least 1 of the 2 data sets. We observe an increased occurrence of methionine oxidation and lysine methylation (as reported also by

Mann et al⁴) and also of deamidation (attributed to boiling under basic condition during sample preparation) and formylation. The much higher occurrence of carbamidomethylation in FrFr is explained by the fact that much less peptides were identified containing C (0.5% in FFPE vs 16.4% in FrFr).