

Additional Variable Modification Search Options in Protein Prospector

Peter R. Baker¹ and Robert J. Chalkley¹

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



Introduction

Users would often like to search for a wider range of variable amino acid modifications than is usually practical in a typical database search program. One way to achieve this is to allow filters to be selectively applied to the modifications. In the latest version of Protein Prospector these include Common, Rare, Max 1..N (where N is the value of the Max Mods parameter) and a motif filter (Fig. 1). Only a single Rare modification can occur in a peptide. The Max options can be used to reduce the value of the max mods parameter for a particular modification. A motif can be specified in the form of a regular expression like those in the MS-Pattern program to further restrict the search space. For example the motif N[notP][SorT] specified as N[^P][ST] reduces the search space by around 80%. Similar filtering options to these are also available in the search engine Byonic.

N- and O-linked glycosylation offer a good test for these search options. A recent study¹ identified 50 different masses of glycosylation modifications ranging from 203 Da to 2543 Da. A further complication with glycosylation analysis is that the higher mass modifications often correspond to the sum of two or more lower mass modifications. However N-linked modifications have an associated motif of N[^P][ST] and many glycosylation sites are annotated in the UniProt database, so the results can be checked.

Methods

Mouse livers were homogenized, then a two-step differential solubilization and centrifugation protocol was followed to prepare a crude nuclear extract that was also enriched in ER, golgi and membrane vesicles. Proteins were digested using trypsin, then glycopeptides were enriched using three rounds of lectin weak affinity chromatography using POROS-WGA. The glycopeptide-enriched tail of the chromatography was further fractionated by high pH reverse phase chromatography, then each fraction was analyzed using an LTQ-Orbitrap Velos, where sequential HCD (measured in the Orbitrap) and ETD (measured in the ion trap) were acquired on each precursor.

The 28275 MSMS spectra were searched using the parameters in Table 1 opposite. Four different filter variants were compared for the glycosylation modifications. The combination 'common' with no motif represents the search format of most search engines. By adding a rare filter in combination with a motif the search time was reduced by a factor of 6 whilst increasing the number of detected proteins by 6% and the number of unique glycopeptides by 23% (Table 2). An example of the reduction in match ambiguity is shown in Fig. 2. The glycopeptides were counted by exporting the results to MS-Viewer and using some recently added sorting and filtering options (Fig. 3).

To investigate the effect of the motif on a mass modification search the proteins identified in the rare motif search were searched looking for mass modifications between -100 and 3500 Da on N instead of the glycosylation modifications. The use of a motif was again found to reduce the search time and increase the number of hits at a given FDR (Table 3).

Conclusions

The new search options substantially reduced search times whilst increasing the number of hits at given FDR threshold. Apart from glycosylation studies these search options could be useful for other modifications, such as hydroxyproline formation in collagens, which have an associated motif of PG, or targeting phosphorylation sites of specific kinases.

References

1. Medzihradzky, K. F., Kaasik, K., Chalkley, R. J. Tissue-specific glycosylation at the glycopeptide level Mol Cell Proteomics (2015) M115.050393

Acknowledgements

This work was supported by grant number NIH NIGMS 8P41GM103481 and the Howard Hughes Medical Institute.

New Variable Modification Options

Modifications are now sorted into categories

Crosslinking
Frequent
Glycosylation
Quant Others
Quant SILAC
Unusual

+ and - button are used to add and remove items from the variable mods menu

Only a single rare modification can occur in a peptide

Offset of the modification site relative to the motif

Up to 6 user defined modifications can be defined

Motifs can be defined for up to 2 amino acids on the mass modification menu

Fig. 1 New Batch-Tag and MS-Tag Search Menu Items

Increasing Hits Whilst Reducing Search Times

Database=SwissProt.2015.4.29.random.concat , taxonomy = MUS MUSCULUS (16718 entries), max missed cleavages=2
Tolerance precursor=15ppm, fragment=0.6 Da
Max variable mods=2
Acetyl (Protein N-term), Acetyl+Oxidation (Protein N-term M), Gln->pyro-Glu (N-term Q), Met-loss (Protein N-term M), Met-loss+Acetyl (Protein N-term M), Oxidation (M)
N-linked Glycosylations (N):
HexNAc, HexNAc2, HexNAc2Fuc, HexNAc2Hex, HexNAc2Hex10, HexNAc2Hex2, HexNAc2Hex2Fuc, HexNAc2Hex3, HexNAc2Hex3Fuc, HexNAc2Hex4, HexNAc2Hex4Fuc, HexNAc2Hex5, HexNAc2Hex5Fuc, HexNAc2Hex6, HexNAc2Hex7, HexNAc2Hex8, HexNAc2Hex9, HexNAc2Hex9Fuc, HexNAc3, HexNAc3Hex3, HexNAc3Hex3Fuc, HexNAc3Hex4, HexNAc3Hex5, HexNAc3Hex6, HexNAc3Hex6Fuc, HexNAc4, HexNAc4Hex3, HexNAc4Hex3Fuc, HexNAc4Hex4, HexNAc4Hex5, HexNAc4Hex5Fuc, HexNAc4Hex5FucSAOx, HexNAc4Hex5FucSAOx2, HexNAc4Hex5SAOx, HexNAc4Hex5SAOx2, HexNAc4Hex5SAOx3, HexNAc4Hex5SAOxSAOxAc, HexNAc5, HexNAc5Hex3, HexNAc5Hex3Fuc, HexNAc5Hex4, HexNAc5Hex4Fuc, HexNAcFuc
O-linked Glycosylations (ST):
HexNAc, HexNAc2, HexNAcHexSA, HexNAcHexSA2, HexNAcHexSAAc, HexNAcHexSAAc2, HexNAcHexSAAcSAOxAc, HexNAcHexSAOx, HexNAcHexSAOx2, HexNAcHexSAOxAc2, HexNAcHexSAOxSAOxAc, HexNAcHexSAOxSAOxAc2, HexNAcHexSAOxSAOxAc3, HexNAcHexSAOxSAOxAc4

Table 1: Database Search Parameters (28275 Spectra)

	motif	time (min)	num proteins	num decoys	num peptides	% FDR	num glycopeptides	unique glycopeptides
rare	N[^P][ST]	11.5	570	6	4075	0.4	1290	503
rare		13	562	6	3832	0.4	1137	458
common	N[^P][ST]	50	561	7	3847	0.6	1089	478
common		72	542	5	3555	0.4	942	410

Table 2 Glycosylation Search Comparisons

Effect of Motif Option On Mass Modification Search

motif	time (min)	num proteins	num decoys	num peptides	% FDR
N[^P][ST]	8	440	6	3673	0.8
	25	417	6	2658	0.7

Table 3 Mass Modification Search Comparisons

Improved Glycopeptide ID

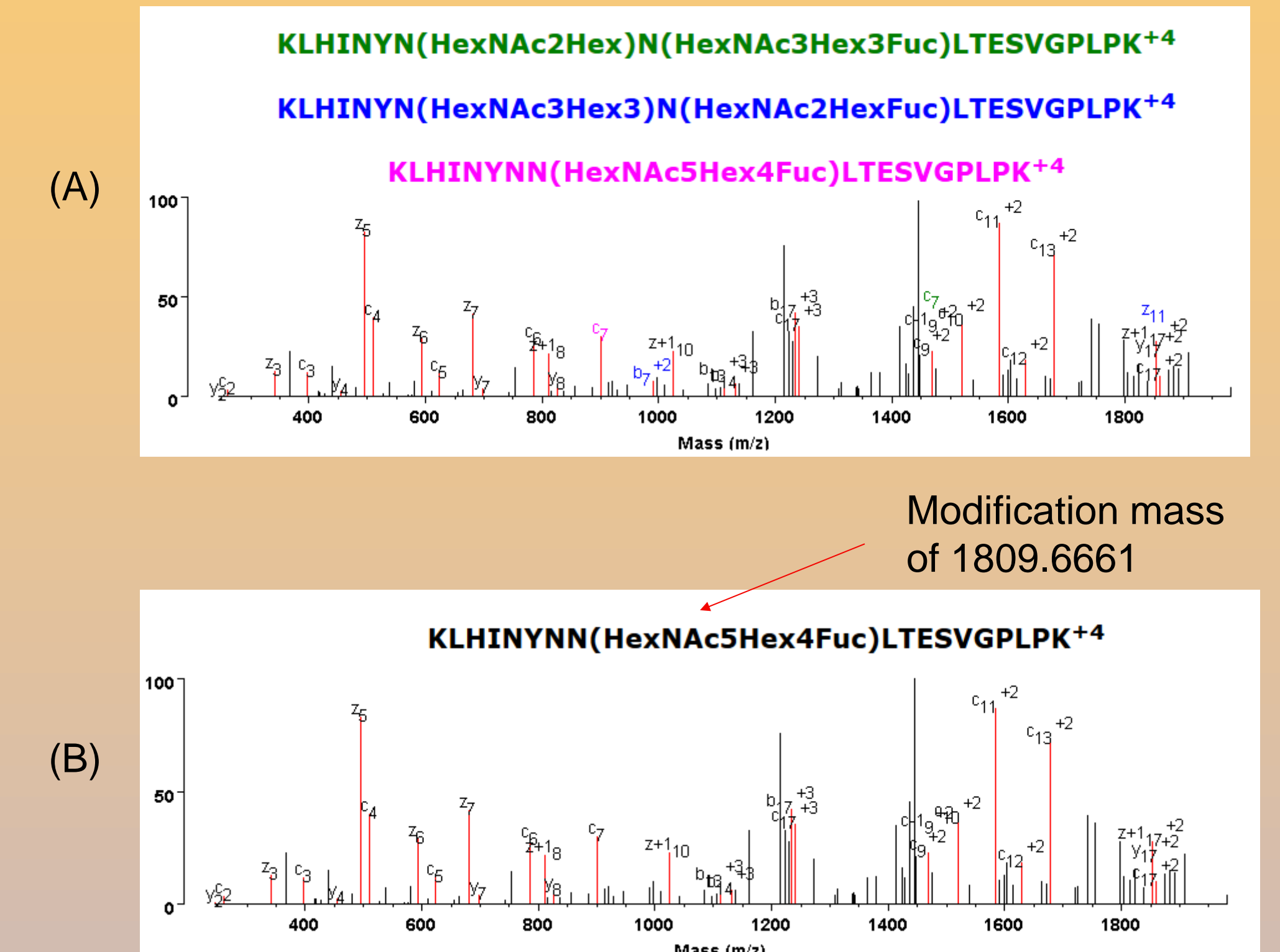


Fig. 2 Filtering Removes Ambiguity

(A) A search with standard parameters reports three potential modification combinations, with ambiguity in terms of modifications. (B) By specifying the glycosylation modifications as rare, and requiring the N-glycosylation to occur in the consensus motif, all ambiguity is removed.

New MS-Viewer Filtering Options

Fig. 3 MS-Viewer Filtering Options Used to Count Glycopeptides

If the results are exported into MS-Viewer the various sorting and filtering options can be used to count glycopeptides. The Report Filtering options can be used to eliminate rows not matching a given criterion. If the remaining results are then sorted then the Remove Replicates option can be used to discard all but the first replicate for up to 5 specified columns.