Increasing Glycopeptide Identifications through Targeted Y Ion Detection using MS-Filter

Robert J. Chalkley, Peter R. Baker

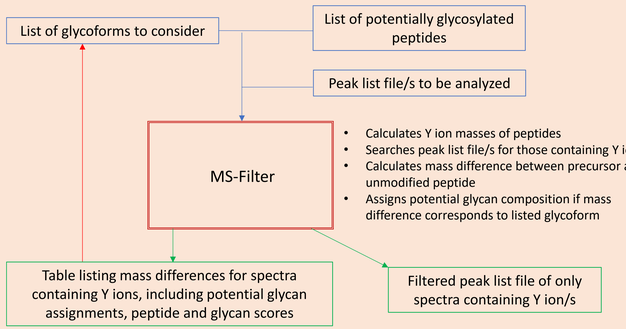
Mass Spectrometry Facility, University of California San Francisco, San Francisco, CA.

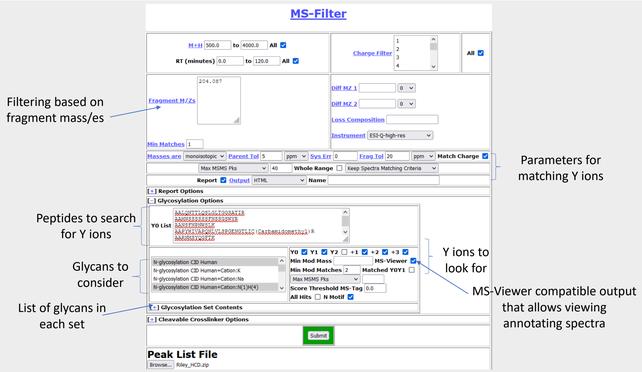
Introduction

Intact glycopeptide analysis using mass spectrometry has become increasingly effective due to hybrid fragmentation protocols such as EThcD and stepped HCD. These have allowed the identification of thousands of glycopeptide spectra from complex mixtures. However, within these datasets there are typically spectra of other glycoforms of the same glycopeptides that are not identified due to either poorer quality fragmentation data or unanticipated modifications. Bioinformatic strategies to uncover these extra glycoforms would provide more comprehensive sample characterization.

On this poster we present developments in the software MS-Filter[1], part of Protein Prospector and freely available on the web, which allows searching peaklists for spectra containing Y ions, then scoring peptide and glycan identifications.

Workflow for Glycopeptide Analysis using MS-Filter





Example Output Displayed in MS-Viewer

Quality of Peptide Match

Quality of Glycan Match

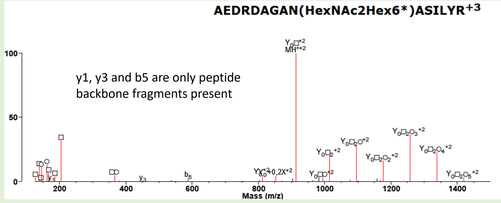
#B ions consistent and inconsistent with glycan assignment

m/z	z	RT	MSMS Info	DB Peptide	Mods	Mod Mass	Error (ppm)	Score	Glyco Score	Num B +	Num B -	HexNAc-C2H6O3 126.0550	HexNAc-CH6O3 138.0550
946.7457	3	36.246	7708	AEDRDAGANASILYR	HexNAc2Hex5*@9	1216.4296	2.38	25.0	21	12	0	103	399
946.7450	3	36.314	7545	AEDRDAGANASILYR	HexNAc2Hex5*@9	1216.4277	1.71	20.8	20	11	0	127	378
946.7460	3	36.190	7728	AEDRDAGANASILYR	HexNAc2Hex5*@9	1216.4305	2.70	19.3	20	11	0	120	384
946.7449	3	36.298	7522	AEDRDAGANASILYR	HexNAc2Hex5*@9	1216.4271	1.50	13.2	20	11	0	157	452
946.4096	3	36.253	6786	AEDRDAGANASILYR	1215.4214@9	1215.4214	0	12.3	-8	10	0	155	426
1000.7644	3	36.023	7687	AEDRDAGANASILYR	HexNAc2Hex6*@9	1378.4856	3.31	6.7	19	9	0	100	357
822.3837	3	36.082	7335	AEDRDAGANASILYR	843.3437@9	843.3437	0	4.9	3	5	0	219	756
887.3997	3	19.936	3972	AEDRDAGANASILYR	1038.3917@9	1038.3917	0	4.0	7	9	0	253	411
887.3994	3	19.949	3663	AEDRDAGANASILYR	1038.3909@9	1038.3909	0	4.0	8	10	0	251	410
670.3316	3	55.738	12813	AEDRDAGANASILYR	387.1873@9	387.1873	0	3.6	1	3	0	623	887
887.4002	3	19.915	3955	AEDRDAGANASILYR	1038.3932@9	1038.3932	0	3.3	8	10	0	233	380
1000.7644	3	36.049	7696	AEDRDAGANASILYR	HexNAc2Hex6*@9	1378.4856	3.31	2.7	19	9	0	94	338
1000.7631	3	36.123	7674	AEDRDAGANASILYR	HexNAc2Hex6*@9	1378.4818	2.03	1.8	17	9	0	58	363

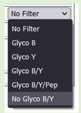
...
(Displays columns for all possible B and Y ions with intensities, if observed)

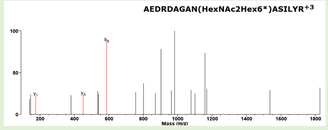
Rescuing Glycopeptide Identifications with Poor Peptide Fragmentation

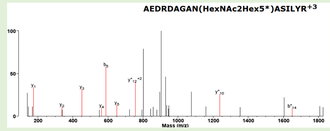
- In the above example the version of the peptide with HexNAc2Hex5 modification was confidently identified by database searching, but none of the HexNAc2Hex6 versions were above the 1% FDR cut-off
  - low peptide score.



- One display option in MS-Viewer is to remove all B/Y ions, allowing easier viewing of peptide evidence







- Comparing the peptide fragments in the HexNAc2Hex6 identification to those observed in the confidently identified HexNAc2Hex5 equivalent, it is clear this is a reliable peptide and glycan identification.

Scoring of Similar Mass Glycan Modifications

- Errors in monoisotopic peak assignment are common for large glycopeptides
- Ammonium adducts are common in glycopeptide datasets enriched by HILIC (using ammonium buffers)
  - Ammonium adduct is +17 Da; which can be confused with +16 Da (e.g. Fuc vs Hex)

Three potential glycan interpretations for the same spectrum

				Error (Da)	Peptide Score	Glycan Score
12511	ANHSLDVSFYFR	HexNAc5Hex4Fuc3*@2	2100.7492	-0.258	16.6	11
12511	ANHSLDVSFYFR	HexNAc5Hex4FucNeuAc*@2	2100.7492	-0.00307	16.6	1
12511	ANHSLDVSFYFR	HexNAc5Hex3Fuc2NeuAc+Cation:N(1)H(4)*@2	2100.7492	-0.261	16.6	1

- Glycan scoring identifies the Fuc3 version (with incorrect monoisotopic mass) as correct answer.

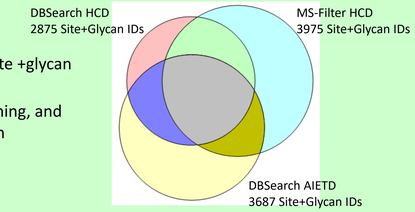
Glycan Scoring

- Based on number of B and Y ions observed that are consistent and inconsistent with glycan assignment
- Penalties for sialic acids and phospho assignments if diagnostic B ions are not observed
- Penalty for Fuc assignment if no Fuc-containing B or Y ions observed
- Scoring effective at correctly ranking assignments, but does not currently have a reliability measure associated with it.

Performance on an Example Dataset

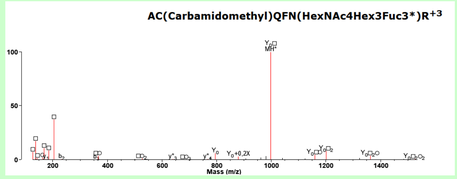
HCD-triggered AI-ETD dataset[2]

- MS-Filter identifies about a third more unique site +glycan IDs compared to database search of HCD data
- HCD database searching, AI-ETD database searching, and MS-Filter all provide complementary information



How Reliable are the Results?

There are many results such as the example to the right, where the peptide identification is probably correct, the mass matches the glycan and there is no contradictory evidence, but there is limited glycan information.



Conclusions

- MS-Filter has been developed to report peptide and glycan scores for glycopeptide assignments initially made through the presence of Y ions from specified glycopeptides
- It is effective at differentiating between isobaric and similar mass glycan assignments.
- It is able to identify more glycoforms of glycopeptides than by database searching alone.
- MS-Viewer allows visualization of peptide and glycan assignments.

Future Developments

- Improved glycan scoring that can better estimate reliability of assignments.

Links and References

Protein Prospector website

MS-Filter

[1] Mol Omics (2020) 16 2 147-155

[2] Nat Commun (2019) 10 1 1311

This work was supported by the Dr Miriam and Sheldon G. Adelson Medical Research Foundation.