

Glycan Scoring Incorporated into Protein Prospector Database Search Results



Robert J. Chalkley, Peter R. Baker

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

Introduction

Several software tools are now available for large-scale analysis of intact glycopeptide spectra. In general, these tools perform well at reliably identifying the peptide portion of the glycopeptide, but are inconsistent at assigning the glycan component. When there are alternative glycan interpretations of similar mass, software needs to correctly consider all the possible compositions, it needs to score which fits best, and decide whether there is sufficient information to call a result reliable.

The first Human Glycoproteomics Initiative Study showed Protein Prospector to be a superior tool for glycopeptide identification, but at that time there was no glycan scoring. In this poster we present incorporated glycan scoring that allows reliable interpretation among possible glycan compositions.

The workflow makes use of MS-Filter¹, 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.

1. Search for glycopeptides by database searching using e.g. Batch-Tag.
2. Input list of potentially glycosylated peptide sequences into MS-Filter.
 - There is a direct link to do this from Batch-Tag
3. Identify additional glycoforms and score between potential glycan assignments.
 - Unexplained mass shifts may correspond to modifications or glycoforms not considered

We show in some cases there can be dramatic differences in the number of glycopeptide spectra that can be identified in comparison to database searching. All data presented on this poster is HCD data from a published analysis of human milk¹.

Example Output Displayed in MS-Viewer

m/z	z	RT	MSMS Info	DB Peptide	Glyco Mod	Mod Mass	Offset	Error (ppm)	Pop Score	Glyco Score	Num B	Num Y	HexNAc-C2H6O3	HexNAc-CH6O3	HexNAc-C2H4O2	HexNAc-C2H4O2
1476.5449	2	17.084	3801	NSGDCPKK	HexNAc4Hex5Fuc2NeuAc*	2059.7360	0	0.362	14.7	23	15	12	0	78	923	43
1476.5449	2	17.084	3801	NSGDCPKK	HexNAc4Hex4NeuAc2+Carbon-N(1)H(4)*	2059.7360	0	-3.44	14.7	8	-15	11	1	78	923	43
984.6995	3	17.902	4033	NSGDCPKK	HexNAc4Hex3Fuc2NeuAc*	2059.7374	0	0.844	4.0	15	12	4	0	123	934	53
984.6995	3	17.902	4033	NSGDCPKK	HexNAc4Hex4NeuAc2+Carbon-N(1)H(4)*	2059.7374	0	-2.96	4.0	3	-12	4	0	123	934	53
984.7007	3	17.194	3834	NSGDCPKK	HexNAc4Hex3Fuc2NeuAc*	2059.7410	0	2.06	13.0	27	33	11	1	11	0	3
984.7007	3	17.190	3832	NSGDCPKK	HexNAc4Hex3Fuc2NeuAc*	2059.7410	0	2.06	6.1	21	15	9	0	116	888	38
984.7007	3	17.190	3832	NSGDCPKK	HexNAc4Hex4NeuAc2+Carbon-N(1)H(4)*	2059.7410	0	-1.74	6.1	6	-15	8	1	116	888	38
984.7007	3	17.194	3834	NSGDCPKK	HexNAc4Hex4NeuAc2+Carbon-N(1)H(4)*	2059.7410	0	-1.74	13.0	6	-33	10	2	11	0	3
994.0321	3	17.219	3840	NSGDCPKK	HexNAc4Hex3Fuc2NeuAc*	2087.7352	0	0	6.5	0	0	0	0	0	0	0
994.0321	3	17.223	3842	NSGDCPKK	HexNAc4Hex3Fuc2NeuAc*	2087.7352	0	0	2.2	0	0	0	0	0	0	0
1033.3851	3	17.082	3800	NSGDCPKK	HexNAc4Hex3Fuc2NeuAc*	2205.7942	0	0.447	1.2	17	15	6	0	115	718	46
1033.3851	3	17.082	3800	NSGDCPKK	HexNAc4Hex3NeuAc2*	2204.7909	1	5.95	1.2	2	-15	5	1	115	718	46
1033.3851	3	17.082	3800	NSGDCPKK	HexNAc2Hex11+Carbon-N(1)H(4)*	2205.7942	0	9.00	1.2	0	-17	3	1	115	718	46
1033.3851	3	17.086	3802	NSGDCPKK	HexNAc4Hex3NeuAc2*	2204.7909	1	5.95	9.9	-6	-33	9	3	12	127	0
1033.3866	3	17.738	3985	NSGDCPKK	HexNAc4Hex5Fuc2NeuAc*	2205.7987	0	1.90	1.6	14	9	5	0	107	892	54
1033.3866	3	17.738	3985	NSGDCPKK	HexNAc4Hex3NeuAc2*	2204.7954	1	7.40	1.6	5	-9	4	1	107	892	54
1042.7185	3	17.383	3887	NSGDCPKK	HexNAc4Hex5Fuc2NeuAc*	2233.7944	0	0	1.0	0	0	0	0	0	0	0

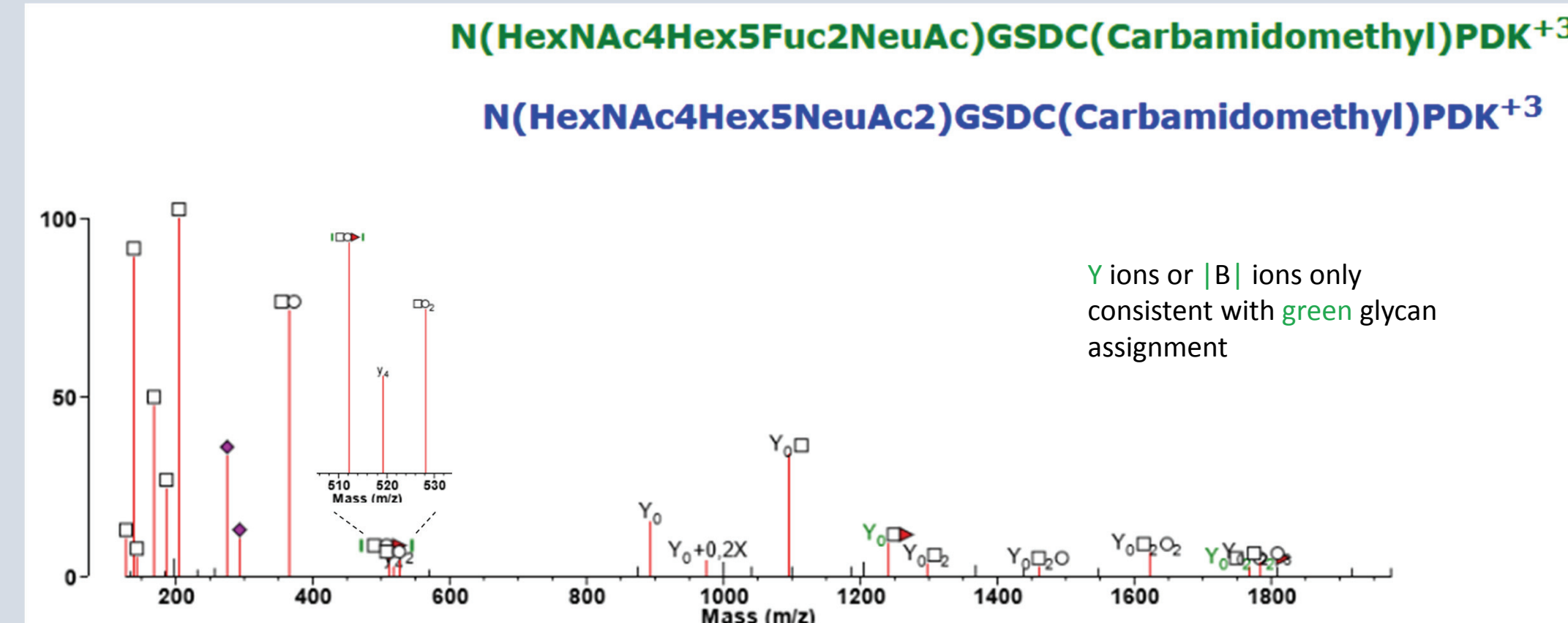
Alternative glycan assignments for same spectrum

Allowing for incorrect monoisotopic peak assignment

Columns report relative intensity of B and Y ions

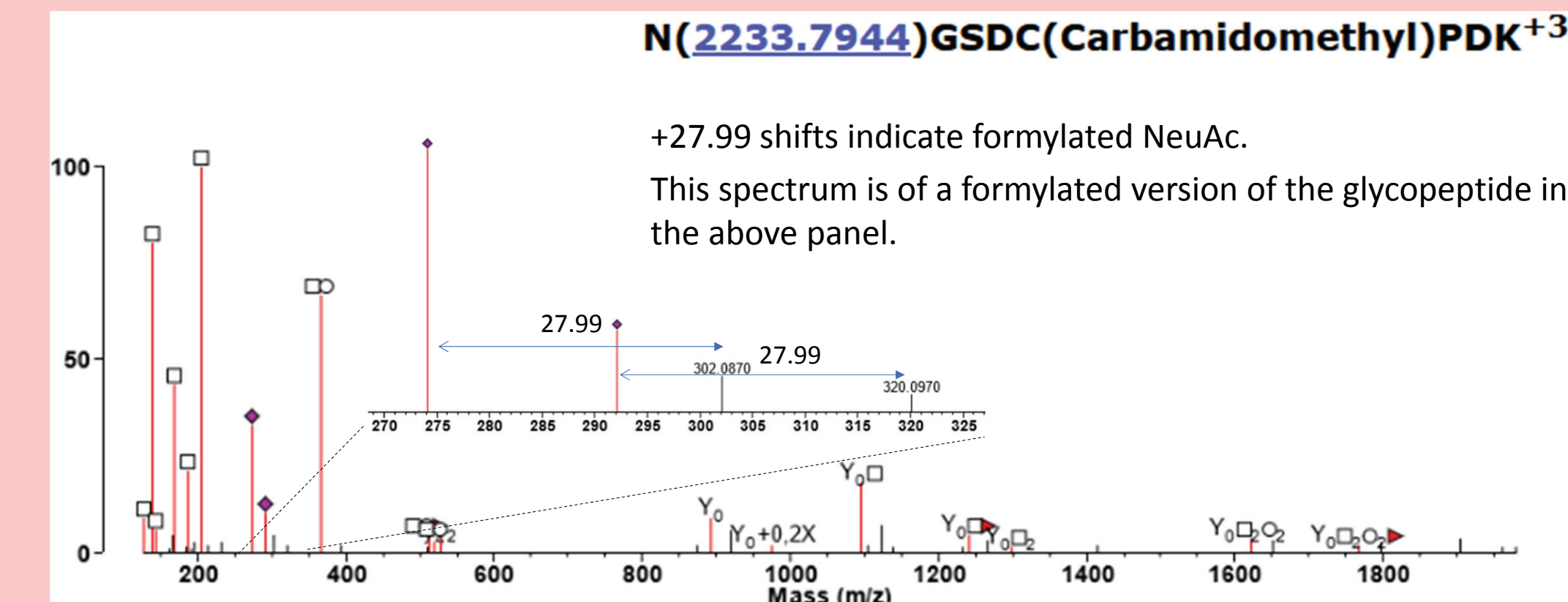
Rescuing Glycopeptide Identifications with Poor Peptide Fragmentation

- In the above example scan 3985 identifies the peptide with either HexNAc4Hex5Fuc2NeuAc2 or HexNAc4Hex5Fuc2NeuAc modification, but with a peptide score of 1.6.
 - Only one peptide backbone fragment (y4) is observed
 - Glycan scoring indicates HexNAc4Hex5Fuc2NeuAc is the correct answer



Identifying Unpredicted Glycoforms

- 2233.7944 Da does not correspond to an expected glycan mass.



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
- Glycan score difference reported when multiple glycan assignments could explain mass.
- Can filter to only report top glycan scoring assignment per spectrum.

Performance in Comparison to Protein Database Searching and Other Software

Human Milk Dataset¹

HCD data from HCD-pd-ETHcD analysis.

Peptide only 1% FDR	2282 Glycopeptide Spectra
pGlyco ²	8582 Glycopeptide Spectra
MS-Fragger Glyco ³	14676 Glycopeptide Spectra (no glycan assessment)
MS-Filter	16265 Glycopeptide Spectra (glycan score threshold of 3)

- Using a single collision energy many glycopeptide IDs have to be based on similarity of fragmentation and elution time relative to a smaller number of higher quality peptide fragmentation spectra.

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.
- It is able to filter results based on relative retention time compared to other glycoforms of the same peptide.
- Its performance compares favorably to other software
- MS-Viewer allows visualization of peptide and glycan assignments.

References

Software is freely available through the web at:

<https://prospector.ucsf.edu>

MS-Filter glycan assignment is currently available; glycan scoring and RT filtering will be in next release.

¹Anal. Chem. (2019) 91 16 10401–10406

²Nat Methods. (2021) 18 12 1515–1523

³Nat Methods. (2020) 17 11 1125–1132

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