GlycoMaster: Glycopeptide High-Throughput Identification and Characterization Software

Bioinformatics Solutions Inc.

University of Waterloo

Lin He[1], Baozhen Shan[2], Lei Xin[2], Gilles A. Lajoie[3], Bin Ma[1]

David R. Cheriton School of Computer Science, University of Waterloo, Waterloo, ON [1]

Bioinformatics Solutions Inc, Waterloo, ON [2]

Department of Biochemistry, University of Western Ontario, London, ON [3]

Overview

We present a software tool, GlycoMaster, for the automated identification of intact glycopeptides from MS/MS spectra through simultaneously searching protein sequence and glycan structure databases. From HCD spectra alone, the software can confidently determine the composition of glycans, together with a list of potential peptides bearing these glycan chains. If using both HCD and ETD spectra, the software can further unambiguously determine the glycopeptide sequences and the glycosylation sites.

Introduction

Glycosylation is one of the most common post-translational modifications (PTMs) and is involved in various diseases. Not only is the fragmentation pattern of a glycopeptide more complex than a non-glycosylated peptide, but also the combination of the protein sequence database and the glycan structure database greatly increases the search space. To meet these challenges, a new software package, GlycoMaster, is developed to confidently identify the glycan compositions and the glycopeptide sequences from MS/MS data through simultaneously searching a protein database and the GlycomeDB database [1]. The MS/MS data analyzed by GlycoMaster can be either HCD spectra or the combination of HCD/ETD spectra.

Methods

Given an MS/MS dataset, a protein sequence database, and a glycan structure database as inputs, the software's workflow includes the following steps:

1. MS/MS spectra are searched against the given protein database to identify a list of proteins through the non-glycosylated peptides existing in the sample.
2. The glycopeptides' MS/MS spectra are selected according to the existence of glycan signature ions (with m/z values 204 and 366) and monosaccharide tags.

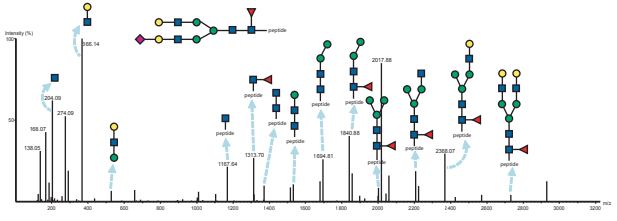

Figure 1. A glycan-spectrum match reported by GlycoMaster. The glycan composition is determined as $(Fuc)_1(Hex)_5(HexNAc)_4(NeuAc)_1$.

3. The glycan structure candidates from the GlycomeDB database are fragmented and matched to each MS/MS spectrum. The glycan composition that best matches a spectrum is reported as the glycan identification.
4. If ETD spectra are available, the glycopeptide sequences are determined by the ETD spectra. Otherwise, a list of possible glycopeptides that both contain N-glycosylation motifs and match the precursor masses are reported.

Data Set

A data set from a human urinary proteome experiment [2] was used to evaluate the performance of GlycoMaster. The sample was preprocessed using lectin affinity enrichment and then analyzed by an LTQ-Orbitrap Velos mass spectrometer (Thermo Fisher Scientific™, Bremen, Germany). The dataset contains 5,008 HCD MS/MS spectra.

Results

Among the 5,008 MS/MS spectra analyzed by our software, 487 have glycan signature ions, and 166 of them were selected as glycopeptide-generated spectra according to our glycan tag de novo sequencing algorithm. These 166 spectra were searched against the GlycomeDB database. 114 spectra were identified with high confidence. Glycopeptide sequences were also searched in the protein list reported by PEAKS 6 database search, 69 out of 114 spectra were identified with glycopeptide sequences. Fig. 1 illustrates a glycan-spectrum match reported by GlycoMaster.

Similar glycan compositions were observed at the same glycosylation site. Table 1 shows such an example: four spectra were identified as the same peptide sequence, LHEITNETFR, but with a slightly different glycan attached. More examples are given in Fig. 2, which was manually annotated based on the protein coverage view of PEAKS 6. Glycopeptides reported by GlycoMaster were annotated in Fig. 2, accompanied by their glycan structures.

Table 1. Similar glycan compositions were observed at the same glycosylation site. Four spectra were identified as the same peptide sequence, LHEITNETFR, but with a slightly different glycan attached.

Scan No.	Precursor m/z	Precursor Charge	Glycan Structure	Score	Glycan Composition
1357	1010.10	3		49.76	$(Fuc)_1(Hex)_5(HexNAc)_4$
1370	961.41	3		40.22	$(Hex)_5(HexNAc)_4$
1535	1107.13	3		23.40	$(Fuc)_1(Hex)_5(HexNAc)_4(NeuAc)_1$
1584	1058.45	3		37.50	$(Hex)_5(HexNAc)_4(NeuAc)_1$

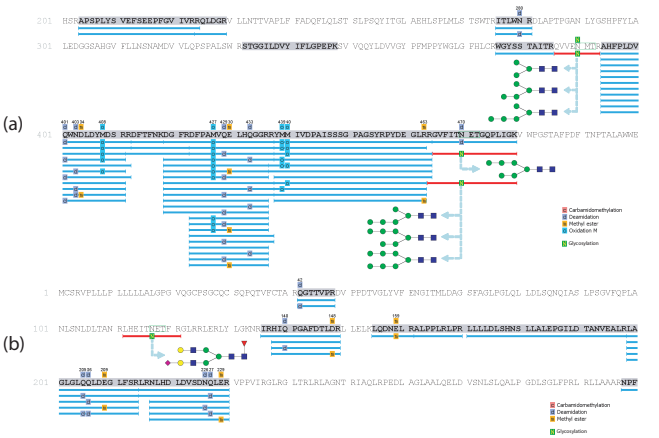

Figure 2. Protein coverage views of two proteins identified from a human urine sample: (a) lysosomal alpha-glucosidase, and (b) vasorin. Glycopeptides, marked with red bars, as well as the glycan structures are reported by GlycoMaster.

Conclusions

GlycoMaster automatically and simultaneously determines the glycan composition and the glycopeptide sequence from an HCD spectrum or an HCD/ETD spectrum pair of an intact glycopeptide. The software is currently available as an online web server at http://www-novo.cs.uwaterloo.ca:8080/GlycoMasterDB.

References

[1] A. Marimuthu, R. O'Meally, R. Chaerkady, Y. Subbannayya, et. al. A Comprehensive Map of the Human Urinary Proteome. Journal of Proteome Research, 10:2734, 2011.

[2] R. Ranzinger, S. Herget, T. Wetter, C. Lieth. GlycomeDB – Integration of Open-Access Carbohydrate Structure Databases. BMC Bioinformatic, 9:384, 2008.

[3] J. Saba, S. Dutta, E. Hemenway, R. Viner. Increasing the Productivity of Glycopeptides Analysis by Using Higher-Energy Collision Dissociation-Accurate Mass-Product-Dependent Electron Transfer Dissociation. International Journal of Proteomics. In Print, 2012.