GlycoP: a software for intact glycopeptide identification and quantification using mass spectrometry-based methods

Lucas Cardoso Lázari[1], Claudia Blanes Angeli[1], Giuseppe Palmisano[1]

[1] USP, Universidade de São Paulo, Department of Parasitology, Institute of Biomedical Sciences, University of São Paulo, São Paulo, Brazil;

Introduction

Protein glycosylation is crucial in many biological processes. Identification of intact glycopeptides allows a complete structural understanding of the micro and microheterogeneity and how its regulations may change in diseased conditions. Accurate glycopeptide identification is vital for the understanding of protein functions, interactions and disease roles, aiding biomarker discovery or improving vaccines design. Several computational platforms are available to characterize the glycosylation profile of single proteins, simple mixture, or complex proteomes. However, most of these solutions are based on high resolution data and are not tailored for specific biotechnology products. Here we developed a software to identify and quantify intact glycopeptides from LC-MS/MS data that can be used for both low- and high-resolution data emphasizing its application in quality control of biotechnology products.

Materials and methods

Human Gonadotropin and Basigin mass spectrometry data was used to develop and test the software. The files were also tested in Byonic software for benchmark purposes. LC-MS/MS data of Gonadotropin was performed in the LTQ-Orbitrap Velos acquired in high resolution in the MS1 (orbitrap) and low resolution (ion trap) in MS2. LC-MS/MS Basigin data was obtained from the article *"Automated N-glycopeptide identification in glycoproteomics"* (Lee et al., 2016), which used a Q-Exactive with high resolution in MS1 and MS2. Computational platform was developed in Python using the PyOpenMS package for spectrum processing.

Results

The GlycoP was developed using different scoring systems, including scores at different levels such as for glycans, peptides and glycopeptides. A clustering algorithm was developed to classify the spectral matches in high, medium and low scoring glycopeptide spectra. These were manually validated to ensure high quality spectra are being selected. A total of 1124 glycopeptide spectral matches were identified for HCD data and compared with the Byonic software. The GlycoP identified a higher number of high scoring spectra confirming its capability for identifying intact glycopeptides.

Conclusions

The GlycoP software has the potential to support the quality control of different biotechnological products. The glycosylation features are extracted and compared between different raw files. We believe that GlycoP can be a useful tool in glycoproteomics research and development.

References

LEE, Ling Y. et al. Toward automated N-glycopeptide identification in glycoproteomics. **Journal of proteome research**, v. 15, n. 10, p. 3904-3915, 2016.

Agradecimentos: This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Centro de Facilidades para a Pesquisa (CEFAP)/Mass-Spectrometry Core Facility (BioMass)