Automatic identification of site-specific glycosylation in proteomics using mass spectrometry and bioinformatics

Jong Shin Yoo

Korea Basic Science Institute, Republic of Korea

E-mail: jongshinyoo@gmail.com

Protein glycosylation, one of the most time-honored posttranslational modifications in proteins, plays vital roles in organic systems through a variety of processes, such as adhesion, signaling via cellular recognition, and response to unusual biological states. However, due to the complexity and heterogeneity of a glycoprotein, contemporary analyses focus in most cases on the identification of both glycosites or the released glycans only. In this study, we have developed MS-based high throughput technique for intact N-glycopeptides analysis, named GlycoProteomeAnalyzer (GPA) for analysis of N-and O-glycosylation in proteomics, which combines tandem Mass Spectrometry (MS) with a database search and algorithmic suite. We created novel scoring algorithms for confident identification of N- and O-glycosylation of proteins with calculation of False Discovery Rate (FDR). In our approach, all amino acid sequence as well as glycosylation website records had been got from the Uniprot database. From the Swiss-Prot accession variety of human protein, our GPA software automatically construct tryptic N- and O-glycopeptide database for the proteins in human plasma sample. It lets in automatic identification of site-specific N- and O-glycopeptides of protein combos using HCD, CID, and ETD MS/MS spectra with GPA-DB from Uniprot with estimated FDR $\leq$ 1%. GPA has been designed to easily handle high-throughput glycoproteomic statistics with a graphical user interface and verified on internet site (https://www.igpa.kr/). It can also be integrated with cloud computing carrier that eliminates the need for local clusters and will increase throughput of data analysis. Glycosylation, a particular enzymatic method in which glycans are attached to proteins or lipids, and is an essential biological method that plays a role in cell signalling, mobile adhesion, and the regulation of biochemical pathways. Of all post-translational modifications (PTMs), glycosylation is one of the most frequently found type. It is believed that more than 50% proteins are glycosylated. The biological functions of glycoproteins are worried many types of organic processes. Therefore, automatic equipment for the identification of glycoproteins and the glycans that are attached to them, become essentially important. For analysis, tandem mass spectrometry (MS/MS) is a popular and environment friendly technique in glycoproteomic due to the fact of its high sensitivity and selectivity.Glycoproteins commonly exist as populations of glycosylated variants (glycoforms) of a single polypeptide. Although the same glycosylation equipment is accessible for all proteins that enter the secretory pathway in a given cell, most glycoproteins emerge with characteristic glycosylation patterns and the glycans at each glycosylation site are heterogeneous. The awareness of equal motifs in distinct glycans lets in a heterogeneous populace of glycoforms to participate in unique organic interactions. This is the most difficult issue for glycan analysis. Two principal kinds of protein glycosylation are known: N-Linked glycans that contain asparagine-X-serine/threonine sequons (N-X-S/T) where X is any amino acid except proline. O-Linked glycans attached to the hydroxyl oxygen of both serine, threonine, tyrosine, hydroxylysine, or hydroxyproline side-chains, or oxygen atoms on lipids such as ceramide symbolize the second kind of modification. Mass spectrometry (MS) has been efficiently used to determine glycan composition and their structures. MS strategies that are presently in use for glycoprotein analysis. These methods can generally be divided into top-down and bottom-up strategies. The determination of the molecular weight of a glycoprotein represents a regular top-down analysis, which affords the most direct method for obtaining records on glycans in a glycoprotein. By calculating the molecular weight variations between the peaks, it is viable to decide the types of glycan modifications on that protein. Such an analysis, however, regularly lacks sensitivity and structure information. Because of this, glycoprotein analyses are often divided into two primary strategies for accumulating glycoprotein information using MS techniques. One entails an evaluation of launched glycans, while the other includes characterizing glycopeptides that are got by proteolytic digestion of the unique glycoprotein.