

GlycoMaster — A software for interpretation of glycopeptides from MS/MS spectra

Baozhen Shan¹, Kaizhong Zhang¹, Bin Ma¹, Cunjie Zhang² and Gilles Lajoie²

¹Department of Computer Science, University of Western Ontario, London, ON, Canada N6A 5B7

²Department of Biochemistry, University of Western Ontario, London, ON, Canada N6A 5C1

Introduction

With progress in proteomics, comes an increasing interest in the importance of glycosylation. The elucidation of glycoprotein structures and functions remains one of the most challenging tasks faced by biochemists and bioanalytical chemists. Tandem mass spectrometry has been recognized as a useful technique for the characterization of the glycan structures. However, manual interpretation of tandem mass spectra of glycoprotein is tedious and time consuming. Instead, computer software programs, STAT (Gaucher et al., 2000), GlycoMod (Cooper et al., 2001), and StrOligo (Ethier et al., 2002), were developed for automated interpretation of MS/MS spectra. These bioinformatics tools are effective, but of limited size of glycan, constrained type of glycoprotein, and human intervention at certain extent. In this poster, we describe a robust, fully-automated computer software program, GlycoMaster, which is designed to determine the glycan structure from variety of MS/MS data of glycopeptides. The structure of a glycan attached to a peptide can be computed if the sequence or mass of the peptide is known.

Method

Scoring Function: In an MS/MS spectrum, each connected subcomponent of the glycan may yield a specific m/z. A matching score can be attributed to a subcomponent according to the peak that matches the subcomponent mass. If no peak matches, a negative score is given. Our software outputs a structure that has the optimal total subcomponent matching scores with the input MS/MS spectrum.

Heuristic Algorithm: GlycoMaster computes the best possible sequence structure among all possible monosaccharide combinations. Analogous approaches have been described, but were computationally inefficient and abandoned. Heuristic dynamic programming technique was used in GlycoMaster. We assume the structures are derived with most probably glycosylation at each step in a network of competing biosynthetic pathways. The software first generates many acceptable small substructures, which are then joined together in a repetitive process to obtain larger and larger suboptimal substructures. Eventually, the (sub)structures reaching the desired mass are output in the order of their scores.

Input and Output: GlycoMaster can read MS/MS spectra in several different formats including lists of experimental masses separating them by spaces or new lines, raw data files or text file with each mass and intensity on a new line. The mass values may be average or monoisotopic. The experimental masses may correspond to glycopeptides or free oligosaccharides. Data from other manufactures can be inputted as text files. For each spectrum, GlycoMaster output a list of sequence structures (5 candidates in default) that can possibly generate the MS/MS spectrum, from the most to the least likely structure.

Preliminary Results

GlycoMaster was tested using ten MS/MS spectra of glycopeptides. The ten glycopeptides were derived from the cationic isozyme peanut peroxidase after tryptic digestion. The MS/MS spectra of the samples were obtained by using a Q-TOF2 in the positive ion ESI MS/MS mode with borosilicate nano tips.

The correctness of the automated interpretation was evaluated by comparing with the structure determined by manual analysis. The software gave correct compositions of all the ten samples. In this initial set of experiments, all the structures are N-linked structures and the number of monomers ranged from 5 to 14. The structures computed by GlycoMaster for eight out of ten spectra are the same as deduced by manual interpretation. The remaining two structures are very similar to the manually

