Scoring Algorithms for characterizing site-specific protein glycosylation

Chuan-Yih Yu [1]; Anoop Mayampurath [1]; Yunli Hu [2]; Yehia Mechref [2]; Haixu Tang [1]
[1]Indiana University, Bloomington, IN; [2]Texas Tech University, Lubbock, TX

Overview

We present computational methods for identifying glycopeptides from their High-energy C-trap Disassociation (HCD) and Collision-Induced Dissociation (CID) tandem mass spectra using scoring algorithm for glycopeptide identification and sequencing algorithm for acquiring glycan sequence.

- HCD scoring
- CID scoring
- Glycan sequencing

Introduction

Glycosylation is one of the most common post-translational modifications (PTMs), and it is associated with several human diseases. Comprehensive glycosylation information comprise identification of glycosylation site, and determination of glycan composition, sequence or structure. We utilize different dissociation methods in mass spectrometry for acquiring glycosylation site and can be used to construct glycan sequence information. HCD produces seven characteristic peaks indicative of mono-, di- or tri-saccharide originating from glycan. CID spectrum contains different part of mono-saccharide fragments of glycopeptide, and this spectrum can be used to construct glycan. Here we report two scoring algorithms introduce a method for glycan sequencing.

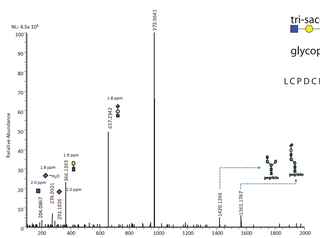
Conclusions

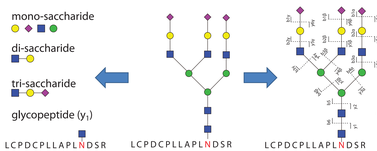
Our tools, GlypID and GlycoSeq, implement scoring algorithm and sequencing method. Manual validations confirmed our preliminary result of the site-specific glycosylation in several glycoproteins. We are currently building a comprehensive framework for glycomics and glycoproteomics, which integrate the computational methods for analyzing HCD, CID and ETD fragment spectra.

Methods

HCD Fragmentation Scoring[1]

- Base on 7 characteristic peaks (Table 1)
- From a binomial distribution
- Compute p-value



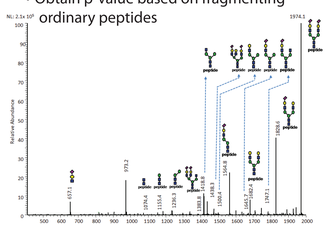


Ion #	m/z	Description
1	138.05	Oxonium Ion @ HexNAc -2 H_2O
2	163.06	Hex
3	204.09	HexNAc
4	274.09	NeuAc – H_2O (water loss)
5	292.08	NeuAc
6	366.14	Hex + HexNAc (di-saccharide combination)
7	657.23	NeuAc + Hex + HexNAc (tri-saccharide combination)

Table 1: List of HCD characteristic peaks

CID Fragmentation Scoring[2,3]

- Count longest path from one peak to next whose inter-peak spacing corresponds to saccharide mass
- Take the longest path as score
- Obtain p-value based on fragmenting ordinary peptides

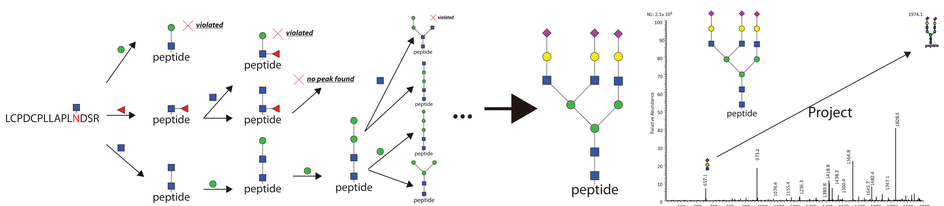


De Novo Glycan Sequencing[4]

- Use CID spectrum
- Start from Y1 (peptide + 1 GlcNAc)
- Use glycan synthetic pathway as growing rules
- Find next peak whose spacing is equal to mono-saccharide, or di-saccharide ***m/z***
- Grow glycan sequence recursively until match with precursor ***m/z***

Heuristic to improve glycan sequencing

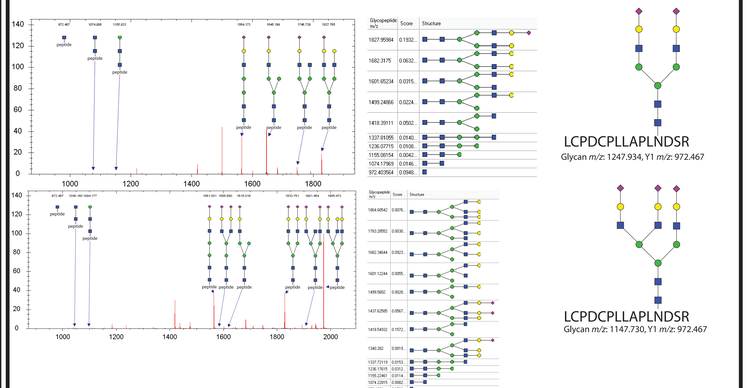
- Add pseudo y-ion by substracting precusor with corresponding b-ion
- Allow one missing fragment
- Use bi-, tri- saccharide to complete the full structure



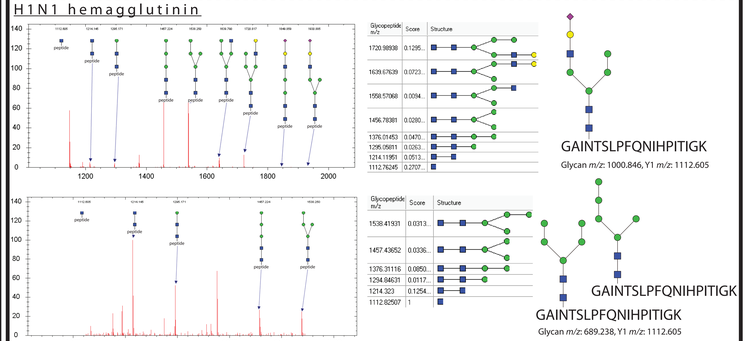
1:Mayampurath et al. Rapid Comm. Mass Spectrom. 2011; 25:2007-2019 | 2:Wu et al. RECOMB : Systems Biology and Computational Proteomics 2007; 4532-96-107
3.Wu et al. Rapid Comm. Mass Spectrometry ; 24 (7) : 965-72 | 4:Tang et al. Bioinformatics., 2005. 21(Suppl 1): p. i431-i439.

Results

Fetuin protein



H1N1 hemagglutinin



This work is funded by National Science Foundation grant number DBI-0642897.