Tandem mass spectrometric analysis of human urinary O-glycopeptides using EThcD activation

Ph.D. dissertation summary

Ádám Pap

Supervisor:

Zsuzsanna Darula Ph.D.

senior research fellow

Biological Research Centre

Laboratory of Proteomics Research

University of Szeged

Faculty of Science and Informatics

Doctoral School of Biology

2021.

Szeged

INTRODUCTION

Glycosylation is among the most common post-translational modifications of proteins. Its roles in living organisms are very diverse. Glycans linked to proteins participate in various immune processes, influence the 3D structure and half-life of proteins. They play a key role in cell-cell adhesion. The two main types of glycosylation are N- and O-glycosylation. In N-glycosylation, the glycan is attached to the protein via the N atom of an Asn side chain, while in O-glycosylation it is attached to the protein through the O atom of a Ser/Thr side chain.

O-glycans are categorized according to the monosaccharide directly attached to the protein. In humans, the most common type of O-glycans are the so-called mucin-type O-glycans in which the monosaccharide directly attached to the protein is a GalNAc. The synthesis of mucin-type O-glycans takes place in the Golgi apparatus. During the synthesis additional monosaccharides are attached to the initiator GalNAc to form the eight core structures of mucin-type O-glycans. These structures may further be elongated with various monosaccharide units (e.g., N-acetyl-lactosamine units, blood group antigens). Of the core structures, core 1 (Galβ1-3GalNAc-S/T) and core 2 (Galβ1-3(GlcNAcβ1-6)GalNAc-S/T) are the most common.

Mass spectrometry is the method of choice for the characterization of complex protein mixtures. Whether the goal is protein identification

or characterization of post-translational modifications, fragmentation data i.e., MS/MS spectra provide the necessary information.

O-glycopeptides are composed of two types of compounds that are chemically different and exhibit different fragmentation characteristics. During data analysis not just the peptide but also the modifying glycan(s) have to be identified. The analysis of polysaccharides is inherently problematic because their isomeric constituents cannot be distinguished by their mass. Similarly, there is no simple mass spectrometric method to characterize the exact position and stereochemistry of a glycosidic bond. Thus, under optimal conditions the following information can be obtained about glycopeptides:

(A) peptide sequence,
(B) monosaccharide composition of the glycan,
(C) the relative position of the glycan-forming monosaccharide units,
(D) site of modification.

Mass spectrometric analysis requires multiple MS/MS experiments with different fragmentation techniques to obtain as much structural information as possible on the two chemically different 'ingredients' that make up an O-glycopeptide.

In *ion trap CID*, fragments are formed via single bond cleavages. These fragments in glycopeptide spectra arise from the cleavage of glycosidic bonds providing structural information about the

modifying glycan. The deglycosylated peptide ion (Y_0) formed during the activation provides information on the molecular mass of the unmodified peptide, and the difference between the precursor ion and the Y_0 ion also indicates the additive mass and thus, the monosaccharide composition of the modifying glycan(s). By considering the biosynthetic pathways, the exact type of the monosaccharide units and the exact structure of the glycan can also be deduced.

During *HCD activation*, in addition to the glycosidic bonds, the peptide bonds may also be cleaved due to the multiple collisions. Therefore, it may be possible to assign the peptide sequence from the detected *b* and *y* sequence ions. Furthermore, the Y_0 ion can also be observed which permits the determination of the additive mass and the monosaccharide composition of the modifying glycan.

However, the site of modification cannot be determined by using any of the above-mentioned activation methods as for both activation techniques the preferred fragmentation event is the gas-phase elimination of the glycan.

ETD activation is a fragmentation technique based on the decomposition of a radical ion formed by electron transfer. During fragmentation, cleavage between the nitrogen atom of the peptide bond and the alpha carbon atom results in *c* and $z^{\bullet}$ sequence ions. Since the amino acid side chains remain intact during fragmentation,

the corresponding sequence ions carry the modifying glycan, hence the site of modification can be determined. However, the ETD spectrum does not provide information on the structure of the glycan.

EThcD activation is a combined fragmentation technique in which ETD is performed first and then HCD on the entire ion set. The complementary HCD step primarily affects the already activated but still intact precursor ions. The ETD step enables the identification of the modified peptide sequence and modification site assignment, while HCD activation provides structural information about the modifying glycan. The advantage of this combined activation technique is that all the above information can be obtained from a single spectrum.

AIMS

The focal point of my PhD research was the evaluation of tandem mass spectrometric data of mucin-type O-glycopeptides enriched from human urine. I had at my disposal the largest mass spectrometric dataset obtained to date by the analysis of intact, tryptic O-glycopeptides. **The aim of the data evaluation was to characterize the fragmentation characteristics of these molecules with a new MS/MS activation technique, EThcD**, and to compare this with other fragmentation methods (ion trap CID, HCD, ETD).

Simultaneously with the analysis of the fragmentation characteristics of O-glycopeptides, the exploration of the O-

glycosylation of urinary proteins is another goal. These goals are inseparable. The knowledge gained in the analytical part of the research fundamentally determines the reliability of biologically relevant information extracted from the recorded data.

METHODS

Sample preparation

Urine samples were trypsinized on a cellulose membrane filter unit. Excess reagents used in the sample preparation can be easily washed away, and the peptides produced after tryptic digestion can be separated from the undigested proteins. A two-round wheat germ agglutinin affinity chromatography was used to enrich the glycopeptides from the resulting peptide mixture. Enrichment of glycopeptides from complex mixtures is absolutely necessary to reduce the non-specific background of non-glycosylated peptides formed during tryptic digestion.

LC-MS/MS analysis

The enriched glycopeptide mixtures were on-line fractionated by reversed-phase liquid chromatography. During the mass spectrometric analysis HCD data-dependent EThcD data acquisition was performed. HCD was recorded for all selected precursors, while EThcD was acquired only for potential glycopeptide candidates, i.e. only if the oxonium ion of N-acetylhexosamine (m/z 204.0867) was among the 20 most abundant fragment ions in the HCD spectrum of the selected precursor ion. EThcD analysis is essential only for the glycopeptides,

and it is more time consuming than HCD measurements. With this setup we optimized the data acquisition process according to our needs.

Data interpretation

Mass spectrometric data were interpreted with database search engines (Byonic, Protein Prospector) and manual data evaluation. We also developed a software (GF-Hunter) to facilitate HCD data analysis.

RESULTS

1.) During EThcD activation larger saccharide oxonium ions may also survive

During our pilot study, glycopeptide mixtures isolated from the urine of three donors were analyzed. Firstly, a general database search was performed with the EThcD data. The O-glycosylation of urinary proteins is less well known, therefore the three most common O-glycan structures reported in human serum were defined. During the evaluation of the search results the presence of larger saccharide oxonium ions in the EThcD spectra were noticed. This phenomenon is due to the low-energy supplemental HCD activation. These ions provide valuable information on whether the peptide is modified with several smaller or one larger O-glycan. A statistical analysis was conducted to answer the question: How often intact oxonium ions corresponding to the most common O-glycans can be detected in the EThcD spectra? It was found that larger oxonium ions such as m/z 948

corresponding to the disialylated core 1 structure or m/z 1313 indicating the presence of the core 2 hexasaccharide gave only a weak signal in most EThcD spectra. Thus, their detection strongly depends on the precursor ion's intensity, but they can be used to confirm or refute glycan assignments.

2.) Expansion of the glycan database using undefined modification database search

The identification rate of the general database search was low relative to the total number of EThcD spectra submitted (~5%). This low identification rate has two major possible reasons. The first is the low efficiency of EThcD activation due to low charge density precursors. The second is that not all O-glycans present in the urine were defined in the search. In order to explore what other O-glycans can modify the urinary proteins in addition to the most common structures described in serum, a so-called unspecified modification database search was performed with the EThcD spectra, where modifications within a certain mass range (0-3000 Da) was allowed only on Ser and Thr residues. A histogram was constructed from the additive masses reported by this search, and the most common additive masses indeed corresponded to different oligosaccharide compositions. A more accurate characterization of the modifications can be achieved by the manual evaluation of the EThcD spectra.

3.) The EThcD technique can be used to determine the relative position of the monosaccharide units of the modifying glycan

During the manual data evaluation, some of the additive masses identified from the unspecified modification search corresponded to O-glycans containing mono- or di-O-acetyl sialic acid and/or disialic acid. The presence of larger saccharide oxonium ions has already indicated that low-energy supplemental HCD activation favors single glycosidic bond cleavages. This observation i.e., the preferred single bond cleavages, permitted the determination of the position of O-acetylated sialic acids and disialic acids within the O-glycan structure. Thus, it can be concluded that the EThcD technique with low-energy supplemental HCD activation facilitates the structural characterization of the modifying O-glycan. **A further benefit of this is that the structural isomers of the same glycopeptide can also be distinguished if these are separated by on-line reversed-phase chromatography, as it was demonstrated in the dissertation with the assignment of an isomer pair. A total of 12 new O-glycan structures were identified on urinary glycoproteins** applying the unspecified modification search and manual data interpretation. Among these, O-glycans containing sialyl-Lewis$^{A/X}$ epitope were also identified, in addition to the previously characterized structures.

4.) The combined analysis of EThcD and HCD data is required

In addition to EThcD data, additional O-glycan structures were identified by considering HCD data. During HCD activation, gas phase elimination of the modifying glycan(s) is the preferred

fragmentation event resulting in the detection of the deglycosylated peptide ion (Y_0) in the HCD spectrum. This can be used to determine not only the molecular weight of the unmodified peptide, but also the additive mass of the glycan(s). Knowing the additive mass, the monosaccharide composition can also be calculated. Thus, new glycoforms of glycopeptides previously assigned from EThcD data can be predicted from the HCD spectra. This hypothesis was tested and confirmed with the most commonly identified peptide (^{342}AVAVTLQSH^{350}, Protein YIPF3) on a few data files, but it also became evident that automatization of the process is required because of the sheer amount of data.

5.) Software development for the identification of new glycoforms from HCD data

In order to exploit the information content of the HCD spectra and to accelerate the prediction of potentially new glycoforms, we have developed a software (GF-Hunter) in cooperation that searches for Y_0-Y_1 ion pairs of previously and reliably identified (e.g., from EThcD spectra) O-glycopeptides in HCD spectra. Due to the complexity of the samples, it is not sufficient to rely on the presence of a single ion (Y_0) during the filtering process. In addition to the Y_0 ion, it is absolutely necessary to look for its Y_1 (peptide + HexNAc) pair in the HCD spectra. Requiring the presence of the ion pair in the same charge state reduces the rate of false positive hits. When the software identifies a spectrum fulfilling the acceptance criteria, it assigns the peptide sequence of the detected ion pair to that HCD spectrum and

calculates the additive mass of the modification using the mass of the precursor ion and the Y_0 ion. If the calculated additive mass matches to an entry in the predefined glycan database, the additive mass is assigned as a glycan, which needs to be validated by manually checking the EThcD pair of the corresponding HCD spectrum. The glycan database built into the software can be iteratively expanded so that previously discovered O-glycans can be identified by the software in subsequent screenings.

6.) Finding new structures

After testing, GF-Hunter was used to screen for new glycoforms using a larger dataset consisting of mass spectrometric data of glycopeptide mixtures enriched from the urine of 10 donors. **The presence of 17 novel glycan structures was confirmed** by manual inspection of the corresponding EThcD data. **These novel O-glycans feature blood group antigens, poly-N-acetyl-lactosamine units, or sulfate.**

SUMMARY

A large and representative mass spectrometric dataset was generated by the analysis of intact human mucin-type O-glycopeptides enriched, not from laboratory cell lines, but from the urine of 10 individuals. Deciphering the data revealed that the glycan repertoire of the urinary glycoproteins is highly diverse. The data interpretation methods presented in the dissertation still rely heavily on manual data evaluation and the inspection of O-glycopeptide identifications delivered by the database search engines. Nevertheless, a **total of 57**

different O-glycans on urinary proteins were identified from the tandem mass spectrometric data of O-glycopeptides. **Our research group was the first who identified these structures on specific proteins and positions, for example, blood group antigens site-specifically on specific secreted proteins. In addition, it was shown that minor differences in isomeric glycan structures may influence the chromatographic behavior of the isomeric glycoforms.**

LIST OF PUBLICATIONS

MTMT identifier: 10058600

Cumulative impact factor: 18.593

Publications on which the dissertation is based:

Pap, A., Klement, E., Hunyadi-Gulyas, E., Darula, Z., & Medzihradszky, K. (2018). Status Report on the High-Throughput Characterization of Complex Intact O-Glycopeptide Mixtures. JOURNAL OF THE AMERICAN SOCIETY FOR MASS SPECTROMETRY, 29(6), 1210–1220. http://doi.org/10.1007/s13361-018-1945-7

IF: 3.255 (2019)

Darula, Z., **Pap, A.,** & Medzihradszky, K. F. (2019). Extended Sialylated O-Glycan Repertoire of Human Urinary Glycoproteins Discovered and Characterized Using Electron-Transfer/Higher-Energy Collision Dissociation. JOURNAL OF PROTEOME RESEARCH, 18(1), 280–291. http://doi.org/10.1021/acs.jproteome.8b00587

IF: 4.074 (2019)

Pap, A., Tasnadi, E., Medzihradszky, K. F., & Darula, Z. (2020). Novel O-linked sialoglycan structures in human urinary glycoproteins. MOLECULAR OMICS, 16(2), 156–164. http://doi.org/10.1039/c9mo00160c

IF: 2.273 (2019)

Other publications:

Pap, A., Medzihradszky, K., & Darula, Z. (2017). Using "spectral families" to assess the reproducibility of glycopeptide enrichment: human serum O-glycosylation revisited. ANALYTICAL AND BIOANALYTICAL CHEMISTRY, 409(2), 539–550. http://doi.org/10.1007/s00216-016-9960-7

IF: 3.637 (2019)

Pap, A., Prakash, A., Medzihradszky, K. F., & Darula, Z. (2018). Assessing the reproducibility of an O-glycopeptide enrichment method with a novel software, Pinnacle. ELECTROPHORESIS, 39(24), 3142–3147. http://doi.org/10.1002/elps.201800223

IF: 3.081 (2019)

Chalkley, R. J., Medzihradszky, K. F., Darula, Z., **Pap, A.,** & Baker, P. R. (2020). The effectiveness of filtering glycopeptide peak list files for Y ions. MOLECULAR OMICS, 16(2), 147–155. http://doi.org/10.1039/c9mo00178f

IF: 2.273 (2019)