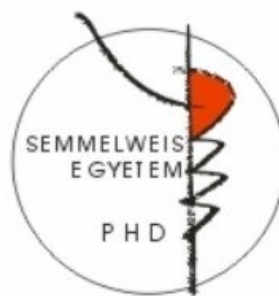# Structural analysis of fucosylated human glycoproteins by tandem mass spectrometry

Thesis booklet

András Ács

Doctoral School of Pharmaceutical Sciences
Semmelweis University

Supervisor: Károly Vékey, D.Sc.
András Telekes, C.Sc.

Official reviewers: Ágnes Alberti-Dér, Ph.D.
Anikó Takátsy, Ph.D.

Head of the Final Examination Committee: Romána Zelkó, D.Sc.

Members of the Final Examination Committee: Viktor Háda, Ph.D.
Pál Riba, Ph.D.

Budapest
2019

I. Introduction

Research over the past decades has shown that proteins encoded by genes may represent a greater diversity of biological functions than the genes themselves. Proteomics is a discipline that studies the structure, function and abundance of proteins. The human genome is made up of about 25,000 genes, while the number of proteins is estimated at over 1,000,000. The structure and function of the proteins may be further modified by various post-translational modifications. Currently, over 200 different post-translational modifications are known, one of the most common of which is glycosylation. An important element of protein structure analysis is the characterization of glycosylation. Modern technological developments make it possible to precisely determine the structure and position of oligosaccharide chains attached to a given protein. During my research I have investigated the position of the fucose substituent on *N*-glycosylated proteins.

There are several widely used methods for studying the glycosylation pattern of a protein. These techniques are routinely used for quality control of biological drugs and for biomarker research. Quantitative changes in fucosylated glycoforms have been described in several biological processes, but the position of fucose has only been determined in a few cases. There is little information weather fucose is always in the same location, or if fucose may be present at different sites is various proteins. The main reason for this is, that identifying the localisation of fucose is not trivial, there are no routine methods for this purpose. Based on the position of fucose, we can distinguish core and antenna fucosylation. During my work I investigated the energy-dependent fragmentation of core and antenna fucosylated model compounds by mass spectrometry. For the identification of core fucosylation, I have chosen PSA (Prostate Specific Antigen) protein, which is known to contain only core fucosylated glycan chains. For characterization of antenna fucosylation I have chosen AGP (alpha-1-acid glycoprotein), which is known to have the fucose substituent only at the antenna. These model compounds were enzymatically digested and the glycopeptides formed were further investigated. The fragmentation processes were characterized by breakdown curves and special formulas developed for this purpose. By analyzing the spectra, I have succeeded to identify diagnostic fragment ions indicative of the types of fucosylation. Using this method I have identified the structure of a rare, bifucosylated glycoform. The method can be easily integrated into common proteomic workflows without the need for additional sample preparation steps. In addition, the technique can be applied to virtually any protein.

II. Objectives

Owing to the development of proteomics and the increased application of therapeutic monoclonal antibodies, the study of glycoproteins has gained importance over the past decade. With the knowledge of the diversity and complexity of sugar side chains, there is a need to define quality parameters for which currently no routinely used methods are available. One such parameter is distinguishing core and antenna fucosylation.

1. The aim of my work was to develop a method based on mass spectrometry that clearly distinguishes core and antenna fucosylation. Based on previous literature data, exclusively core fucosylated PSA and exclusively antenna fucosylated AGP glycoproteins were selected as model compounds for the studies.

2. The first objective was to prepare glycopeptides from the model proteins, than study and characterize their tandem mass spectra. The selected glycopeptides are present in complex protein digests, so their analysis requires separation prior to tandem mass spectrometry. For this reason they were analyzed using nano-HPLC-MS/MS.

3. In order to obtain detailed structural information on the glycopeptides, we planned to characterize them using energy dependent MS/MS. Collision energies were studied in the range of 30% to 150% of the standard collision energy used for proteomics.

4. Energy dependence was interpreted by using breakdown curves. The aim was to identify diagnostic fragment ions and specific energy ranges that will differentiate core and antenna fucosylation.

5. A further aim of my research was to determine the structure of rare, multiply fucosylated AGP glycoforms.

III. Methods

III.1. Enzymatic digestion

From lyophilized AGP and PSA standards, 1 nmol of material was dissolved in 30 µl of water + 5% methanol and used for digestion. For the enzymatic digestion I have used the protocol previously developed by our research group for the digestion of small protein amounts. For protein solubilization and reduction of disulfide bridges, samples were incubated with 5 µl of 0.5% Rapigest and 2 µl of 200 mM DTT (1,4-dithiothreitol) at 60 ° C for 30 min. Subsequently, 2.5 µl of 200 mM IAA (iodoacetamide) and 5 µl of 200 mM NH_4HCO_3 were added and incubated at room temperature for 30 minutes. The trypsin enzyme, when used alone, often fails to cleave the peptide chain alongside the lysine amino acid (so-called missed cleavage). To eliminate this phenomenon, mixture of Lys-C and trypsin enzyme (Lys-C/Trypsin) was added for pre-digestion in a 1:100 enzyme:protein ratio, and incubated for 1 hour at 37 ° C. Lys-C cleaves more efficiently lysine than trypsin, thus its use improves digestion efficiency. Subsequently, trypsin was added in a volume of 1 µl to a 1:25 enzyme:protein ratio and incubated for another 2 hours at 37 ° C. To inactivate the enzymes and to stop digestion I have used 1 µl FA (formic acid). In case of the PSA standard, Lys-C/Trypsin and trypsin digestion resulted in a glycopeptide consisting of two amino acids (NK, monoisotopic mass 260.148) which, due to its low molecular weight, was not suitable for further study. Digestion with Arg-C resulted in the formation of a 9 amino acid glycopeptide (NKSVILLGR, monoisotopic mass 998.624) which can be well investigated in further experiments. Therefore, in the case of PSA standard, the above mentioned protocol was modified by instead of using Lys-C/Trypsin and trypsin, 1 µl of Arg-C enzyme was added for digestion at 1:25 enzyme:protein ratio. For optimal function of Arg-C, 1 µl of 200 mM DTT and 1 µl of 15 mM $CaCl_2$ (calcium chloride) were added to the sample before the enzyme was added and incubated for 10 minutes at room temperature.

III.2. Nano LC-MS/MS

Liquid chromatography was performed on an Ultimate 3000 nanoRSLC system (Dionex, Sunnyvale, CA, USA). An Acclaim PepMap100 C-18 trap column (100 μm × 20 mm; Thermo Scientific, Sunnyvale, CA, USA) was used to desalt the samples and an Acquity UPLC M-Class Peptide BEH C18 column (1.7 μm, 130 Å, 75 μm × 250 mm, Waters, Milford, MA, United States) was applied to separate the peptides. During the separation, a gradient elution of 60 or 90 minutes was used at 48°C, at a flow rate of 300 nl/min, eluent A was water+0.1% FA and eluent B was acetonitrile+0.1% FA. The liquid chromatography system was coupled to a Maxis II ETD Q-TOF mass spectrometer equipped with a CaptiveSpray nanoBooster ion source (Bruker Daltonics, Bremen, Germany).

III.3. Glycosylation analysis

MS1 spectra were recorded in the 150-3000 m/z mass range at 5 Hz. Collision-induced dissociation (CID) was performed on selected abundant precursor ions (>25000 counts/second) at 4 Hz and on low intensity ions (>5000 counts/second) at 1 Hz. The tetraantennary AGP glycopeptides appear in the spectra only with low abundance, therefore I have set the frequency of MS1 measurements to 3 Hz, and the frequency of MS/MS spectra to 1 Hz in the case of intensive ions. The "standard collision energy" was determined according to the instrument manufacturer's recommendations based on the isolation mass range width, the isolation *m/z* and the charge state of the ion. When examining the energy dependence, I have fragmented the given precursor ion with a specified percentage of this energy, in the range of 30% to 150%. Raw data were evaluated using Compass DataAnalysis 4.3 (Bruker Daltonics, Bremen, Germany). The intensity of the precursor ion and the fragment ions were determined by the program. The intensity of a given ion was determined based on the area of its most abundant isotope.

IV. Results

(1) I have developed a method based on tandem mass spectrometry that clearly distinguishes core and antenna fucosylation of glycopeptides.

As a first step, I have selected model glycopeptides for the characterization of core and antenna fucosylation. These should be of high intensity, in order to be able to get reliable energy dependent tandem mass spectra. Based on our previous results and literature data, the selected glycopeptides were exclusively either core or antenna fucosylated. For the characterization of core fucosylation I have chosen the PSA-derived glycopeptide NKSVILLGR-N4H5S2F1, and for the antenna fucosylation, the NEEYNK-N4H5S2F1 glycopeptide was used. Energy-dependent MS/MS spectra of selected glycopeptides were used to determine the composition of fragment ions.

(2) Diagnostic fragment ions were identified in the MS/MS spectra, which characterize core and antenna fucosylation.

Antenna fucosylation is clearly confirmed by the presence of the $NHSF^+$ ion, which is easily detectable at low collision energy. In addition, the complementary $[M-NHSF]^{2+}$ ion is also observed. At low collision energy, the presence of core fucosylation can be indirectly deduced from the absence of $NHSF^+$ and $[M-NHSF]^{2+}$ ions. The location of the fucose can also be identified at higher collision energies. In such a case, the ratios of the $[M-NHS-NHSF]^+/[M-NHS-NHS]^+$ or $[peptide + NF]^+ / [peptide + N]^+$ ion pairs formed by sequential fragmentation processes should be examined.

(3) I have developed mathematical formulas to characterize the degree of antenna fucosylation, based on the relative abundance of diagnostic fragment ions.

I have developed three different equations, based on various fragment ion ratios. These equations are suitable for quantifying the amount of antenna (and core) fucosylation. The practically most useful of these expressions is $F_1 = [M-NHSF]^{2+} / ([M-NHS]^{2+} + [M-NHSF]^{2+})$. Note that partly due to fucose migration and partly to the interference multi-step fragmentation processes, there is a threshold for applicability: Determining less than 2% core fucose contribution (in predominantly antenna fucosylated species) or less than 14% antenna fucose contribution (in predominantly core fucosylated species) is unreliable.

(4) The energy dependence of fragmentation processes were characterized by the breakdown curve method.

Breakdown curves can be used to illustrate energy dependence of each fragmentation mechanism. It was found that neutral fucose loss requires lower activation than the predominant charge separation process. Core fucose is more strongly bound than antenna fucose; and the latter is easier to fragmentation and to migrate.

(5) Optimal collision energy for the characterizing fucose position was established.

The optimal collision energy should be determined individually for each glycopeptide. The optimum is, where the survival yield is greater than 50%. In such a case mostly one-step fragmentation processes occur, and these are easy to interpret. Under such conditions the diagnostic ions are abundant, can be reliably identified, and the chemical noise is very small.

(6) I have identified a rare bifucosylated glycoform in a genetic variant of AGP (A1AG2_HUMAN).

I have determined the structure of the glycopeptide of ENGTVSR-N6H7S4F2 composition derived from AGP. It is a tetraantennary bifucosylated compound (see Figure 1. below), and both fucose substituents are located on the antenna.
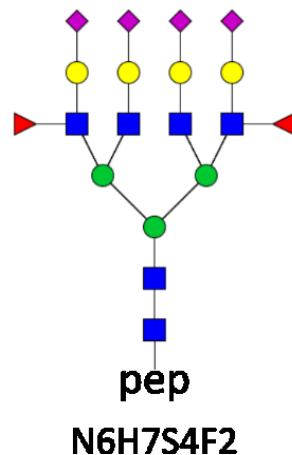

N6H7S4F2

Figure 1.   Structure of the tetraantennary ENGTVSR-N6H7S4F2 glycopeptide

V. Conclusions

During my research I have dealt with structural characterization of fucosylation and the structural identification of glycoforms. The importance of fucosylation has been described in a variety of biological and physiological processes, and with the spread of biologics in the pharmaceutical industry, its importance is increasing. There is currently no widely accepted analytical method for determining the position of fucose. In the dissertation I present a technique where the position of fucose can be determined from low collision energy MS/MS spectra. It is also possible to draw qualitative conclusions from routinely used high collision energy spectra and to compare different samples even using high collision energy spectra. The great advantage of this method is that it does not require derivatization and can easily be incorporated into the conventional proteomic workflow. During development of the method I have investigated *N*-glycopeptides. However, there is no theoretical obstacle to the extension of the method to the analysis of *O*-glycopeptides. It is known from prior literature that identifying the position of fucose may complicated by fucose migration. In the dissertation I show that fucose migration is drastically decreased at low collision energy, and under such condition it is only a minor problem. In the case of low-energy CID fragmentation, when the molecular ion is the most intense peak in the spectrum (e.g. survival yield above 50%), fucose migration and other interfering processes can be minimized. I recommend using 3+ charged glycopeptides for fragmentation assays as they provide the most informative MS/MS spectra. Note that the present investigations were performed on a QTOF type mass spectrometer. Due to the longer residence times, in instruments with ion trap analyzers the probability of fucose migration will increase. A further advantage of low collision energy CID is that MS/MS spectra are very clear, there is no chemical noise and the diagnostic ions are among the most intense spectral peaks. At low collision energy, multiply fucosylated glycoforms can also be identified and the location of fucose substituents can be determined. Detailed characterization of glycan structures contributes to the development of quality control of biological medicinal products and can be an important new tool for biomarker research.

VI. Bibliography of the candidate's publications

VI.1. Publications related to the thesis

1. Ács A, Ozohanics O, Vékey K, Drahos L, Turiák L. (2018) Distinguishing Core and Antenna Fucosylated Glycopeptides Based on Low-Energy Tandem Mass Spectra. Anal Chem, 90: 12776-12782.

IF: 6.350

2. Ács A, Turiák L, Révész Á, Vékey K, Drahos L. (2019) Identification of bifucosylated glycoforms using low-energy CID spectra. J. Mass Spectrom.

IF: 2.267

VI.2. Publications not related to the thesis

1.   Turiak L, Toth G, Ozohanics O, Revesz A, Acs A, Vekey K, Zaia J, Drahos L. (2018) Sensitive method for glycosaminoglycan analysis of tissue sections. J Chromatogr A, 1544: 41-48.

IF: 3.858

2.   Gogl G, Biri-Kovacs B, Poti AL, Vadaszi H, Szeder B, Bodor A, Schlosser G, Acs A, Turiak L, Buday L, Alexa A, Nyitray L, Remenyi A. (2018) Dynamic control of RSK complexes by phosphoswitch-based regulation. Febs J, 285: 46-71.

IF: 4.739

3.   Kovacs AF, Lang O, Turiak L, Acs A, Kohidai L, Fekete N, Alasztics B, Meszaros T, Buzas EI, Rigo J, Jr., Pallinger E. (2018) The impact of circulating preeclampsia-associated extracellular vesicles on the migratory activity and phenotype of THP-1 monocytic cells. Sci Rep, 8: 5426.

IF: 4.011

4.   Turiak L, Ozohanics O, Toth G, Acs A, Revesz A, Vekey K, Telekes A, Drahos L. (2019) High sensitivity proteomics of prostate cancer tissue microarrays to discriminate between healthy and cancerous tissue. J Proteomics, 197: 82-91.

IF: 3.537

5.   Kovács ÁF, Fekete N, Turiák L, Ács A, Kőhidai L, Buzás EI, Pállinger É. (2019) Unravelling the Role of Trophoblastic-Derived Extracellular Vesicles in Regulatory T Cell Differentiation. Int. J. Mol. Sci., 20: 3457.

IF: 4.183

6.   Turiak L, Sugar S, Acs A, Toth G, Gomory A, Telekes A, Vekey K, Drahos L. (2019) Site-specific N-glycosylation of HeLa cell glycoproteins. Sci Rep, 9: 14822.

IF: 4.011

7.   Müller A, Langó T, Turiák L, Ács A, Várady G, Kucsma N, Drahos L, Tusnády GE. (2019) Covalently modified carboxyl side chains on cell surface leads to a novel method toward topology analysis of transmembrane proteins. Sci Rep, 9: 15729.

IF: 4.011