Glycosylation analysis of blood plasma proteins

Thesis booklet

Eszter Tóth

Doctoral School of Pharmaceutical Sciences
Semmelweis University

Supervisor: Károly Vékey DSc

Official reviewers: Borbála Dalmadiné Kiss PhD
Anikó Takátsy PhD

Head of the Final Examination Committee: Tamás Török DSc
Members of the Final Examination Committee: Éva Szökő DSc
László Lelik CSc

Budapest
2016

Introduction

The proteome is defined as the entirety of all protein variants of an organism. The goal of proteomics is study of structure and function of proteins. Proteomics focuses on the qualitative and quantitative analysis of proteins. Furthermore protein-protein interactions, interactions of proteins and their environment (carbohydrates, lipids, nucleic acids, drugs), modification of proteins, protein maps and mechanisms of different physiological and pathological processes are also important research areas of proteomics.

Post-translational modifications yield most of the high diversity of proteins and they take part in fulfilling each protein's special role. Glycosylation is one of the most important post-translational modification of proteins in mammals and more than 50% of proteins are glycosylated. Structural and functional characterization of glycosylated proteins is the goal of glycoproteomics. During the process of glycosylation oligosaccharide chains (glycans) are bounded to the glycosylation sites, which are specific amino acids. Various type of glycans can be bound to the same glycosylation site, this is called microheterogeneity, and different glycoforms are the result of microheterogeneity.

The most important roles of glycosylation are stabilization of the structure of proteins, cell-cell communications and interactions (adhesion, endocytosis, molecular recognition), protein interactions (receptor activation) and glycoproteins are antigens in the immune system. Glycosylation influences the mechanism of action of therapeutic proteins, their pharmacokinetic properties and stability. Moreover biomarker properties of glycoproteins are very important in the diagnosis of various diseases.

Identification of proteins in cells, in tissues and in organs and analysis of special protein properties, among them post-translational modifications is very difficult. Co-operation of different research areas: biochemistry, molecular biology, separation sciences, analytical chemistry and bioinformatics and continuous development of methods are significant points of the glycosylation studies.

Nowadays analysis of glycoproteins is a great challenge, and applicable techniques need to be developed. There is a huge diversity of blood plasma proteins, great number of glycoforms in case of each glycoprotein, and several orders of magnitude difference in the concentration range of various protein types. Therefore reducing the complexity of

samples is always mandatory, and more dimension of separation steps are required before mass spectrometric analysis in every case.

Aims

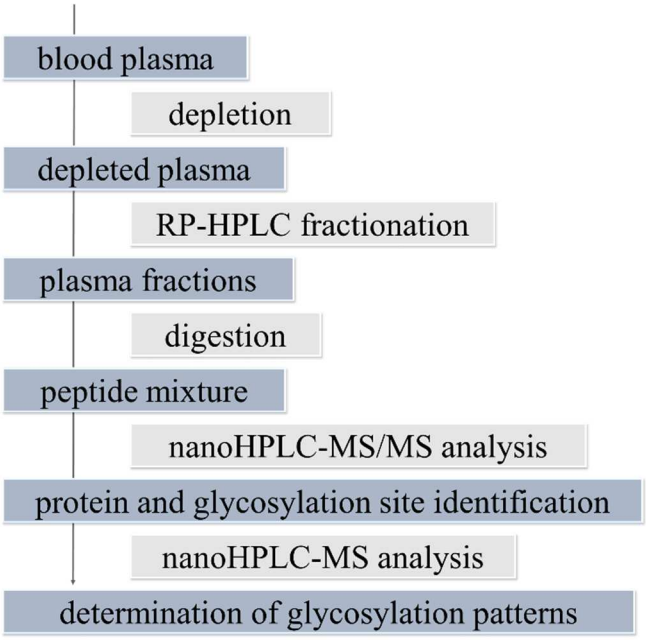
During my PhD work, my goal was to determine the site-specific glycosylation pattern of plasma proteins.

The first step was the development and optimization of a suitable method for determination of plasma protein glycosylation patterns. To this end, it was my goal: to develop a depletion and fractionation method for lowering complexity of blood plasma samples; optimization of the subsequent nano-HPLC-MS(/MS) measurements and improvement of their reproducibility.

After the development and optimization of the methods I sought to determine site-specific glycosylation pattern of multiple proteins, first healthy people, followed by samples of patients treated with radiation therapy. My goal was to investigate the effects of ionizing radiation on glycosylation and characterize occurring changes.

Methods

To determine site-specific glycosylation pattern of blood plasma proteins a multi-step sample preparation, analysis and data evaluation method was optimized and used. The entire workflow is shown below:



Albumin and immunoglobulin G were removed from the blood plasma samples by manual elution on a prepacked affinity chromatography column. The resulting depleted plasma was fractionated by reverse phase chromatography using analytical-scale HPLC column and UV detection. Proteins in each fraction were reduced, alkylated and cleaved into tryptic peptides and glycopeptides.

The composition of the peptide mixture was analyzed by nano-HPLC-MS(/MS). I performed data-dependent MS/MS measurements in order to determine the protein content of the fractions. A full cycle consists of an MS spectrum and MS/MS spectra of the three most intense ions. Identification of glycosylation sites and major glycoforms was done by data-dependent MS/MS measurements using different instrument settings (higher *m/z* range and lower energy). Identification of the low-concentration glycoforms and quantitative analysis was carried out using extended dynamic range MS measurements.

The evaluation of the measured data was partly carried out using commercial and partly using in-house developed software. Identification of the protein content of the fractions was performed using ProteinLynx Global Server with Mascot Server and human

SwissProt protein sequence database. I identified the glycosylation sites and the intensive glycoforms by manual MS/MS spectra evaluation and the GlycoMiner software. The identification of lower intensity glycoforms was achieved by manually searching the theoretical *m/z* values of each glycoform in the MS spectra. I accepted only those peaks as true identification, for which charge state, isotopic pattern and HPLC retention time were as expected. For not clearly identifiable glycoforms, on selected components I recorded individual MS/MS spectra, and glycoforms were identified by evaluating these spectra.

The intensity of glycoforms (relative concentration) was determined using GlycoPattern software. For further measurements, GlycoPattern was used not only to determine the intensity values, but also used to identify the glycoforms. In order to identify as many glycopeptides as possible and filter out false positives, I optimized search parameters then partly automated evaluation with these optimized set of search parameters.

Results

(1) I developed and optimized a sample preparation method consisting of the removal of high quantity proteins followed by the fractionation of plasma by reverse phase chromatography. I proved that this method greatly reduces the complexity of the blood plasma samples, enabling not only the analysis of individual proteins, but of the various glycoforms present in small concentrations. This method made it possible to determine the detailed, site-specific glycosylation patterns of proteins. In addition, it has been proved that analytical HPLC columns (4.6 and 2.1 mm in diameter) are suitable for the production of protein isolate sufficient to determine glycosylation patterns of several plasma proteins. [1]

(2) I proved that using the developed method, different glycoforms of the same protein are not significantly separated from each other during the reverse-phase HPLC. As a result, the glycosylation pattern of a protein can be determined from a single fraction. Furthermore, I found that passage through the reversed phase column does not distort glycosylation patterns, that is, the distribution of glycoforms does not change. This is very different from many other protein isolation methods, which separate the glycoforms from each other, thus greatly complicate the analysis of glycosylation patterns. [1]

(3) Using simple statistical methods, I characterized the reproducibility of several days' long nano-HPLC-MS measurement batches. I found that the standard deviation varies greatly depending on the intensity of the component, and increases significantly in the case of components close to the detection limit. In addition, I found that reproducibility for nano-HPLC is much worse than usual for HPLC-MS; for intense peaks it is around 15%. Based on the tests runs I noticed that the poor reproducibility is caused not just by chance variation, but the intensity of each peak significantly changes with time (over several days). I have shown that the intensity of various components changes over time in varying degrees (and different directions). [2]

(4) I found that changes in ion intensity over several days can be well described by a fourth-order polynomial. I developed a method for correcting the described time dependent change. For the example presented in the thesis I found that the standard deviation could be reduced from 15% to 8% using the developed method. The correction will not only increase the reproducibility of the measurements, but also reduce systematic errors as well. [2]

(5) After optimizing sample preparation, measurement and data evaluation methods, I determined the site-specific glycosylation pattern of several abundant plasma proteins. The resulting patterns were more detailed than can be found in the literature, although a more stringent acceptance criteria was used. I also characterized the site-specific glycosylation patterns for proteins not previously described in the literature. The number of glycoforms identified in blood plasma reaches the number of glycoforms usually only observed for pure protein standards available in large quantities, which means that additional purification or isolation steps at protein level are no longer necessary.

(6) I carried out the first human study on the effect of ionizing radiation on glycosylation of blood plasma proteins. Blood samples were taken prior to radiation treatment, during treatment and 1-1.5 months after the end of treatment. From these samples site-specific glycosylation patterns of 7 proteins were determined. I established that irradiation treatment results in changes in protein concentrations as well as in glycosylation patterns. [3]

(7) I found that changes in glycosylation patterns persist a long time after radiation treatment. The majority of radiotherapy-induced changes were still detectable 1-1.5 months after the end of treatment, although their level was lower than during the irradiation. This shows that treatment changes glycosylation for a long time. The glycosylation of proteins is slowly returning to the state before irradiation, but it may take several months. [3]

Conclusions

My results confirm the hypothesis that reverse phase separation of proteins is mostly governed by the properties of the peptide backbone and glycosylation to have little impact. This is a great advantage, as reverse phase fractionation allows the determination of the glycosylation pattern from one fraction. The same effect is probably not unique to glycosylation, but many other post-translational modifications as well. Initial results demonstrate that it is the case with phosphorylation, although detailed investigations were carried out only for glycosylation.

The correction method based on fourth-order polynomial fitting greatly reduces the standard deviation of long mass spectrometry based experiments. The method was tested for identical and slightly different samples, where small differences had to be determined. The polynomial correction is generally usable for long-term measurement series, so we expect it will spread widely. The pharmaceutical industry is also interested in the described method.

The obtained site-specific glycosylation patterns of plasma proteins are more detailed than those found in the literature, and in some cases can be considered novel. The newly identified glycoforms can bring us closer to understanding the biological significance of proteins and a deeper understanding of their operation. Study of the thus identified glycoforms in various diseases can lead to new, more selective biomarkers. In our laboratory research is extended in this direction.

Ionizing radiation affects glycosylation of multiple proteins and these changes persist long term. Since the glycosylation is an essential element in many immunological processes and also in communication between cells, the detected changes can reveal the effectiveness or side-effects of radiation therapy. The biomedical implications of these results are studied by our polish colleagues.

List of publications

Publications related to the thesis

[1] **Toth E,** Ozohanics O, Bobaly B, Gomory A, Jeko A, Drahos L and Vekey K. (2014) HPLC enrichment/isolation of proteins for post-translational modification studies from complex mixtures. J Pharm Biomed Anal, 98:393-400.
IF (2014): 2.979
10.1016/j.jpba.2014.06.025

[2] **Toth E,** Hever H, Ozohanics O, Telekes A, Vekey K and Drahos L. (2015) Simple correction improving long-term reproducibility of HPLC–MS. J Mass Spectrom, 50:1130-1135.
IF (2014): 2.379
10.1002/jms.3629

[3] **Toth E,** Vekey K, Ozohanics O, Jeko A, Dominczyk I, Widlak P and Drahos L. (2016) Changes of protein glycosylation in the course of radiotherapy. J Pharm Biomed Anal, 118:380-386.
IF (2014): 2.979
10.1016/j.jpba.2015.11.010

Publications not related to the thesis

Vekey K, Ozohanics O, **Toth E,** Jeko A, Revesz A, Krenyacz J and Drahos L. (2013) Fragmentation characteristics of glycopeptides. Int J Mass Spectrom, 345-347:71-79.
IF (2013): 2.227
10.1016/j.ijms.2012.08.031

Bobaly B, **Toth E,** Drahos L, Zsila F, Visy J, Fekete J and Vekey K. (2014) Influence of acid-induced conformational variability on protein separation in reversed phase high performance liquid chromatography. J Chromatogr A, 1325:155-162.
IF (2014): 4.169
10.1016/j.chroma.2013.12.022

Krajsovszky G, **Toth E** and Ludanyi K. (2014) Tandem mass spectrometric study of annelation isomers of the novel thieno[3',2':4,5]pyrido[2,3-d]pyridazine ring system. ARKIVOC, 5:158-169.
IF (2014): 1.165
10.3998/ark.5550190.0015.500