Ph.D. Theses

Oliver Ozohanics

Development and Application of Bioinformatic and Proteomic Methods for Glycoproteomics

Supervisor:

László Drahos, Ph.D.

Institute of Organic Chemistry, Research Centre for Natural Sciences, Hungarian Academy of Sciences

Chemistry Ph.D. School

Ph.D. school leader: Prof. Dr. György Inzelt

Synthetic Chemistry, Material Sciences and Biomolecular Chemistry Program

Program leader: Prof. Dr. András Perczel

Eötvös Loránd University, Faculty Natural Sciences

Budapest, 2012

Introduction

After mapping of the human genome, researchers' attention turned towards proteins. Proteomics and biomolecular mass spectrometry (MS), the latter being one of the prime analytical methods used in protein studies, are among the most dynamically developing research fields today. Developments in instrumentation and bioinformatics resulted in routine use of mass spectrometry for the analysis of biological samples, in particular their protein content. At present structure characterization of proteins, including their post-translational modifications became feasible, just as their quantitation in complex mixtures.

Glycosylation is one of the most frequent post-translational modifications of proteins. It has a key role in several biological processes, like cell-cell communication and immune response. Changes in these processes closely relate to important diseases. A sound knowledge of the structure and function of protein glycosylation may help understanding various biological processes and the underlying cause for various diseases as well. Recognizing changes in the oligosaccharide chains of glycoproteins may have diagnostic significance as well. This is suggested also by the fact that accepted tumor markers are mostly glycoproteins.

Compared to genomics or proteomics, our knowledge of glycosylation is rudimentary. The main reason for this is that structure characterization of glycoproteins is a major analytical challenge. Part of the problem lies in the heterogeneity of glycoproteins: Typically there are many different isoforms present, differing in the structure of the corresponding oligosaccharides (often called glycoforms). Enrichment is often necessary before analysis. Even so, low concentration limits glycoform analysis mostly to chromatography and to mass spectrometry. Further development of enrichments and analytical methodologies are still needed before breakthrough in glycochemistry may be achieved.

Aims

Characterization of glycosylation used to be a very long, multi-step procedure, even when the glycoprotein was available in a relatively large amount. We wanted to develop a workflow for glycoprotein analysis in the biomedical field. This usually involves analysis of tens or hundreds of samples, so relatively high throughput was necessary. Biological samples are complex mixtures, so sample preparation and glycoprotein enrichment was also needed. Mass spectrometry is a key analytical method for glycoprotein analysis, so we based our approach on this technique.

My work was focused towards structural characterization of N-glycosylated proteins. First, tandem mass spectrometric (MS/MS) fragmentation of glycopeptides was needed to be understood in detail. The next step was to automate data analysis (spectra evaluation) by development of an algorithm and its software implementation. Having identified glycoforms based on automatic analysis of MS/MS spectra, their relative amounts had to be established (the glycosylation pattern). A further aim was the indentification of minor glycoforms. These two objectives were solved based on HPLC-MS analysis, supported by another novel algorithm and software. Last but not least, these steps were needed to be integrated into a single workflow.

Following method developments, I also wanted to illustrate capabilities of the developed workflow in practical applications, including identifying biomarkers for some cardiovascular diseases.

Experimental

Analysis of glycosylation pattern of proteins

During my experiments, I determined the glycosylation pattern of standard proteins and proteins isolated from clinical blood plasma samples. Analysis starts with tryptic proteolysis of the glycoprotein (or glycoprotein mixture). The resulting peptide and glycopeptide mixture was analyzed using nano-UHPLC-MS and –MS/MS methods. Data analysis of MS/MS measurements was done using the developed GlycoMiner software, while single stage MS data were analyzed using the GlycoPattern software, specifically developed for these purposes. The figure below illustrates the applied workflow:

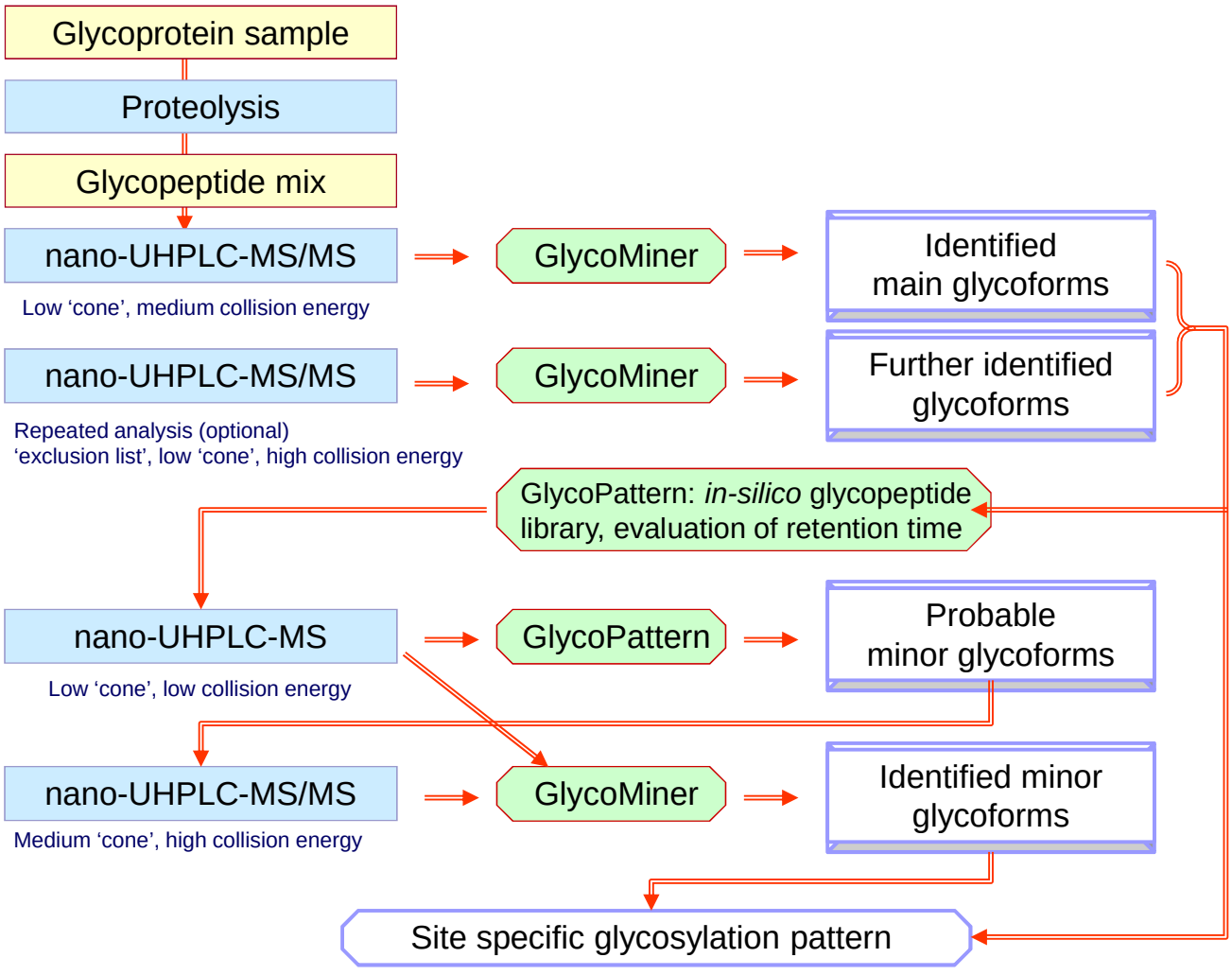

Fig. 1: Workflow for determination of site specific glycosylation

Results and Discussion

During my PhD work I developed the workflow shown above, including the necessary softwares and algorithms. I used the workflow for analysis of prospective atherosclerosis biomarkers and for analyzing the glycosylation pattern of blood plasma proteins. I have also developed a method to enable comparison of glycoprotein and glycopeptide enrichment. The main results and thesis of my research are the following:

1) I developed a nano-UHPLC-MS(MS) based workflow (figure 1.), which enables site specific analysis of protein glycosylation. The most important features of the workflow are that: i) it can be used to analyze glycoprotein mixtures (and not only isolated glycoproteins); ii) it is sufficiently sensitive to analyze low sample amounts (1-10 pmol), therefore applicable for biomedical purposes; iii) it is of "medium" throughput (analysis of hundreds of samples is feasible). The workflow is based on and integrated series of measurements and data analysis. Proteins present in the sample are identified by tandem mass spectrometry. Tandem mass spectrometry is also used to determine structure/composition of glycopeptide isoforms, i.e. site specific glycoprotein structures. Relative quantitation of glycopeptides is based on a third mass spectrometric measurement, yielding the site specific glycosylation pattern. These require generating an *in-silico* database of possible glycopeptides and estimation of their chromatographic retention times.

2) The most challenging part of the previous workflow is the automated structure (composition) determination of glycopeptides based on tandem mass spectra. For this purpose I have developed an algorithm and implemented it in software named GlycoMiner. Tests indicate that glycopeptide identification is very accurate; GlycoMiner has on average 0.1% false positive and 0.1% false negative error rate. A practical advantage that GlycoMiner is capable for analyzing low quality, noisy spectra as well – a common case in proteomics. The software can be freely downloaded from the website of RCNS (http://www.chemres.hu/ms/glycominer). It has become useful for a wide community, as has been downloaded by 150 potential users.

3) I have applied the developed workflow to analyze site specific glycosylation pattern of glycoprotein standards and glycoproteins from blood plasma of individual persons. A major advantage of our workflow that it does not require isolated (purified) glycoproteins, but complex mixtures (like plasma) may also be used to identify glycosylation pattern of a given

component. Glycosylation pattern of several proteins like haptoglobin, antitrypsin, serotransferrin, AGP etc. have been determined from blood plasma.

4) I characterized three methods for glycopeptide and glycoprotein enrichment (fenilboronic acid affinity, wheat germ agglutinin affinity and anion exchange). I have developed a numerical index to characterize glycopeptide content of the fractions. I showed that glycopeptides can be more efficiently enriched, than glycoproteins. I also showed that among the three methods compared that based on fenilboronic acid based yields the best results.

5) The developed workflow was used to detect atherosclerosis biomarker candidates. I determined the site specific glycosylation of AGP in different samples and compared the results of samples from healthy and diseased people with atherosclerosis and aneurysm. Aneurysm samples were used as a positive control. Results show that the site specific glycosylation pattern of AGP is a promising biomarker candidate for atherosclerosis diagnosis, with a good predictive value of 89%. This confirms that the developed sample preparation and analytical techniques, and the GlycoMiner software have practical significance as well. Using these tools we were able to detect over 100 AGP glycoforms using a small sample amount only (ca 10 pmol AGP). The determined pattern may be useful for diagnostic applications distinguishing healthy and diseased individuals, and to monitor progress of disease or treatment.

Publications

Publications related to the Ph.D. thesis

1. Ozohanics, O., Krenyacz, J., Ludanyi, K., Pollreisz, F., Vekey, K., and Drahos, L. (2008) GlycoMiner: a new software tool to elucidate glycopeptide composition. *Rapid Communications in Mass Spectrometry* 22, 3245-3254. IF: 2.846

2. Ozohanics, O., Turiak, L., Drahos, L., and Vekey, K. (2012) Comparison of glycopeptide/glycoprotein enrichment techniques. *Rapid communications in mass spectrometry* 26, 215-217. IF: 2.846

3. Ozohanics, O., Turiák, L., Puerta, A., Drahos, L., and Vékey, K. (2012) HPLC-MS methodology for analyzing site-specific glycosylation patterns. *Journal of Chromatography A*. IF: 4.194 (submitted)

Other publications

1. Turiak, L., Ozohanics, O., Marino, F., Drahos, L., and Vekey, K. (2011) Digestion protocol for small protein amounts for nano-HPLC-MS(MS) analysis. *Journal of Proteomics* 74, 942-947. IF: 5.074

2. Turiak, L., Misjak, P., Szabo, T. G., Aradi, B., Paloczi, K., Ozohanics, O., Drahos, L., Kittel, A., Falus, A., Buzas, E. I., and Vekey, K. (2011) Proteomic characterization of thymocyte-derived microvesicles and apoptotic bodies in BALB/c mice. *Journal of Proteomics* 74, 2025-2033. IF: 5.074

3. Ambrus, A., Torocsik, B., Tretter, L., Ozohanics, O., and Adam-Vizi, V. (2011) Stimulation of reactive oxygen species generation by disease-causing mutations of lipoamide dehydrogenase. *Human Molecular Genetics* 20, 2984-2995. IF: 8.058

4. Zsoldos-Mady, V., Ozohanics, O., Csampai, A., Kudar, V., Frigyes, D., and Sohar, P. (2009) Ferrocenyl pyrazolines: Preparation, structure, redox properties and DFT study on regioselective ring-closure. *Journal of Organometallic Chemistry* 694, 4185-4195. IF: 2.205

5. Budai, L., Ozohanics, O., Ludanyi, K., Drahos, L., Kremmer, T., Krenyacz, J., and Vekey, K. (2009) Investigation of genetic variants of alpha-1 acid glycoprotein by ultra-performance liquid chromatography-mass spectrometry. *Analytical and Bioanalytical Chemistry* 393, 991-998. IF: 3.841

6. Budai, L., Pollreisz, F., Ozohanics, O., Ludanyi, K., Drahos, L., and Vekey, K. (2008) Analysis of complex oligosaccharides using graphitized carbon liquid chromatography/mass spectrometry. *European Journal of Mass Spectrometry* 14, 419-422. IF: 1.34

Presentations held on international conferences

1. Ozohanics, O., Krenyácz, J., Pollreisz, F., Vékey, K., Drahos, L.: *Automatic identification of saccharides and glycopeptides in MS/MS*, 25th Informal Meeting on Mass Spectrometry; Nyíregyháza-Sóstó, 2007. May 6-10. (oral)

2. Ozohanics, O., Krenyácz, J., Budai, L., Ludányi, K., Kremmer, T., Vékey, K., Drahos, L.: *Analysis of genetic variants of α-1 acid glycoprotein in cancer*, Second Central and Eastern European Proteomic Conference; Jena, 2008. October 12-15. (plenary lecture)

3. Ozohanics, O., Krenyácz, J., Budai, L., Ludányi, K., Kremmer, T., Vékey, K., Drahos, L.: *Az alfa-1 savas glikoprotein genetikai variánsainak vizsgálata*, Elválasztástudományi Vándorgyűlés 2008; Sárvár, 2008. November 5-7. (oral)

4. Ozohanics, O., Drahos, L., Vékey, K.: *Glycosylation, mass spectrometry and informatics*, 3rd Central and Eastern European Proteomics Conference; Budapest, 2009. October 6-9. (oral)

5. Ozohanics, O., Memboeuf, A., Indelicato, S., Drahos, L., Vékey, K.: *Size dependence of fragmentation – predictable or mysterious?*, 28th Informal Meeting on Mass Spectrometry; Kőszeg, 2010. May 2-6. (oral)

6. Ozohanics, O., Turiák, L., Lengyel, Á., Vékey, K., Drahos, L.,:*Glycosylation Patterns and Glycopeptide Biomarkers*, 5th Central and Eastern European Proteomic Conference; Prága, 2011. September 19-22. (oral)

7. Ozohanics, O., Turiák, L., Lengyel, Á., Vékey, K., Drahos, L.: *A Challenge for nanoUPLC-MS/MS: Determination of Glycosylation Patterns*, HPLC 2011; Budapest, 2011. June 19-23. (oral)

8. Ozohanics, O., Pollreisz, F., Gráf, L., Vékey, K.: *Intact protein analysis of crayfish trypsin variants*, 27th Informal Meeting on Mass Spectrometry; Retz, 2009. May 3-6. (poster)

9. Ozohanics, O., Vékey, K., Drahos, L.: *Kinetic characterization of the Golgi protein glycosylation pathway*, 3rd EuPA Congress Clinical Proteomics; Stockholm, 2009. June 14-17. (poster)

10. Turiák, L., Ozohanics, O., Buzás, E., Vékey, K.: *Mass spectrometric analysis of T-cell derived membrane vesicle proteins*, 3rd Central and Eastern European Proteomics Conference; Budapest, 2009. October 6-9. (poster)

11. Turiák, L., Ozohanics, O., Misják, P., Buzás, E., Vékey, K.: *Determination and quantification of proteins originating from activated T-cell derived membrane vesicles by mass spectrometry*, 25th International Symposium on Microscale BioSeparations MSB 2010; Prága, 2010. March 21-25. (poster)

12. Ozohanics, O., Vékey, K., Drahos, L.: *N-glycosylation kinetics model*, 4th Summer School on Mass Spectrometry in Biotechnology and Medicine; Dubrovnik, 2010. July 4-10. (poster)

13. Ozohanics, O., Turiák, L., Drahos, L., Vékey, K.: *Glycoprotein and glycopeptide enrichment methods on human plasma samples,* 4th European Conference on Chemistry for Life Sciences; Budapest, 2011. August 31 - September 3. (poster)