Glycomics based biomarkers of the rate of aging : development and applications of high-throughput N-glycan analysis

Ruhaak, L.R.

English summary

Introduction

Glycosylation is the enzymatic addition of oligosaccharides (also known as glycans) to proteins and lipids. More than 50% of human proteins are glycosylated and proper glycosylation is essential for the survival of most organisms as the glycans have important functions in biological processes such as: protein folding, protein clearance, cell adhesion, receptor binding and receptor activation.

As glycans are often branched structures, where monosaccharides may be linked in several different ways, protein glycosylation may be very diverse. Moreover, it is a dynamic equilibrium: within an individual, the glycan signature is highly reproducible in a given physiological state, however, when the physiological state changes, e.g. due to aging or disease, the glycan pattern can change dramatically. Therefore, protein glycosylation is an important group of potential biomarkers. Several types of glycosylation have been described in humans; however, the studies in this thesis are focused on N-glycans, where the oligosaccharide structure is attached to asparagine residues of the protein.

To evaluate the potential of glycosylation patterns as biomarkers, it would be most favorable to analyze cellular body materials, as this is the site of action of most glycoproteins. It is, however, difficult to obtain such material; therefore bodyfluids, which may reflect the cellular processes, are analyzed, similarly to the proteomics and metabolomics fields.

Over the years several strategies have been developed for the analysis of N-glycans, however, a drawback of these methods is that their suitability for high-throughput analysis is limited. The methods may be based on intact protein analysis, glycopeptide analysis or glycan analysis. Strategies for high-throughput analysis developed in this thesis mainly focus on glycan analysis as such a strategy is more widely applicable, but glycopeptide analysis has also been applied. While glycopeptide analysis

has the advantage that the site of attachment of the N-glycan is known, it is difficult to employ on complex plasma samples.

N-glycan analysis is usually performed by enzymatic release of the glycans from the proteins; the glycans may be labeled prior to analysis. Labeling usually increased the ionization efficiency of the glycans in mass spectrometric analyses or introduces a fluorescent or UV-absorbing group to allow their analysis.

Markers for longevity which reflect the health condition and predict healthy aging would be valuable in aging research. Such markers, which indicate biological age of individuals instead of calendar age, have so far hardly been identified. The Leiden Longevity Study (LLS) was especially designed to search for biomarkers that reflect familial longevity and predict healthy aging. The cohort consists of 420 Caucasian families of exceptional longevity. Next to long-lived siblings (n = 960), also their offspring (n = 1710) and the partners of the offspring (n = 761) were recruited, to allow case-control studies. The study design is illustrated in Figure 1-9.

The aim of the studies described in this thesis is to search for changes in glycosylation that are associated with familial longevity in the Leiden Longevity Study. As the LLS comprises several thousands of samples, there existed a need for high-throughput analysis techniques to allow large scale glycan analysis. Thus, development of such techniques –especially at the level of sample preparation- which are also applicable for other study cohorts, has been an essential part of the studies described in this thesis.

Research described in this thesis

In chapter 2 a 96 well plate-based high-throughput procedure for the rapid preparation of 2-aminobenzoic acid (2-AA)-labeled N-glycans from human plasma is described. During this procedure, N-glycans are released from glycoproteins and subsequently labeled with 2-AA without prior purification. A HILIC-based solid phase extraction (SPE) method is then applied to isolate the 2-AA-labeled N-glycans, which

can be analyzed by MALDI-TOF-MS, HPLC with fluorescence detection (FL), and CE-MS. Up to 4 times 96 human plasma samples can be handled in parallel, which, together with the versatility of the 2-AA label, makes this procedure very attractive for glycomics analysis of larger sample cohorts.

Labeling of N-glycans (as performed in Chapter 2) is usually performed by reductive amination with a fluorophore containing a primary amine to allow fluorescence detection. For reductive amination, the amine group of the label reacts with the reducing-end aldehyde group of the oligosaccharide to form a Schiff base, which is reduced to a secondary amine. To reduce the Schiff base, a reducing agent is used. In glycan labeling the almost exclusively used reductant is sodium cyanoborohydride, however, during the reaction this is converted to the toxic HCN gas. In chapter 3, the use of 2-picoline-borane as a non-toxic alternative to sodium cyanoborohydride as the reducing agent for oligosaccharides is described.

While the use of HPLC-FL for the generation of glycosylation patterns as described in Chapter 2 is effective, it is not very efficient. Therefore, a need existed for further separation-based techniques that allow quantitative analysis. An optimized sample preparation method for N-glycan-profiling by multiplexed capillary gel electrophoresis with laser-induced fluorescence detection (CGE-LIF) is described in Chapter 4, enabling high-throughput glycosylation analysis. Glycans are released enzymatically from denatured plasma glycoproteins, and labeled with APTS using 2-picoline-borane. Reaction conditions are optimized for a high labeling efficiency, short handling times, and only limited loss of sialic acids. Samples are then subjected to hydrophilic interaction chromatography (HILIC) purification at the 96 well plate format prior to analysis by CGE-LIF using a 48-capillary DNA sequencer. The method was found to be robust and suitable for high-throughput glycan analysis. Even though the method comprises two overnight incubations, 96 samples can be analyzed with an overall labor allocation time of 2.5 hours.
The second part of the thesis comprises mainly applications of the methods described above to the Leiden Longevity Study.

In Chapter 5, the method comprising 2-AA labeling and HILIC-HPLC-FL analysis as described in Chapter 2 was applied to investigate whether glycans derived from the total glycoprotein pool in plasma mark familial longevity and distinguish healthy from unhealthy aging in the LLS. Two N-glycan features (LC-7 and LC-8) could be identified to be more abundant in plasma of the offspring of long-lived individuals as compared to their partners as controls. These results were not confounded by the altered lipid status or glucose homeostasis of the offspring. The glycan features could be associated with CRP levels, indicating that glycosylation probably also marks health status in this cohort. Furthermore, a decrease in levels of LC-8 was associated with the occurrence of myocardial infarction, indicating that plasma glycosylation patterns do not only mark familial longevity, but may also reflect healthy aging. Further studies are needed to confirm these results.

While Chapter 5 focuses on N-glycosylation patterns of total plasma, Chapters 6 and 7 are dedicated to N-glycosylation patterns of specific plasma glycoproteins. It has been shown previously that the N-glycosylation pattern of human immunoglobulin G (IgG) is age-dependent. In Chapter 6 it is assessed whether N-linked glycans on IgG also reflect early features of human longevity. IgG subclass specific glycosylation patterns were obtained from 1967 participants in the LLS by MALDI-TOF-MS analysis of tryptic IgG-Fc glycopeptides. It could be corroborated that the degree of galactosylation of IgG decreased with increasing age. Moreover, for the galactosylated glycoforms the incidence of bisecting GlcNAc increased as a function of age .In younger participants (<60 years of age), but not in the older age group (>60 years) decreased levels of non-galactosylated glycoforms containing a bisecting GlcNAc reflected early features of longevity. Therefore, it may be concluded that IgG glycoforms mark calendar age at all ages and the propensity for longevity before middle age.

The use of large-scale immuno-affinity capturing combined with high-throughput N-glycan analysis for targeted glycan based biomarker discovery is described in Chapter 7. Alpha-1-antitrypsin and immunoglobulin A are enriched from human plasma in

a bead-based procedure in a 96-well microtitration plate based platform. The strategy has been applied to 2395 plasma samples from the Leiden Longevity Study. AAT-glycosylation patterns are associated with calendar age and differ between females and males. Moreover, several AAT-glycans could be associated with physiological parameters marking cardiovascular and metabolic diseases. Two non-fucosylated glycans were shown to be positively related to the incidence of myocardial infarction. While several parameters could be related to AAT-glycosylation features, the IgA glycosylation patterns seem to be almost unaffected.

Conclusions

As the final chapter (Chapter 8) of this thesis, a general discussion, which places the individual chapters into context, is included. Several changes in N-glycosylation patterns associated with familial longevity were observed in this thesis. In chapter 5, two glycan features originating from total plasma proteins could be associated with familial longevity. Moreover, three IgG glycoforms, all non-or mono-sialylated and containing a bisecting GlcNAc, were found to reflect familial longevity in chapter 6. These altered patterns did not interfere with previously found features reflecting familial longevity and may thus be regarded novel. The predictive value of the observed glycan-markers is, however, rather low. Therefore, it has to be concluded that multiple markers are needed for the prediction of complex phenotypes like familial longevity.

The present thesis describes the development of several approaches for high-throughput N-glycosylation analysis at the level of the total plasma proteome as well as at the level of individual proteins. These techniques were then applied to a biomedical study involving a cohort size which is typical for such studies, and which indeed needs the throughput which was technologically realized. It is anticipated that the developed methodology will be of considerable value to other fields of biomedical research.