SHORT THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY (PHD)

Glycoanalytical identification of early differential diagnostic markers by capillary electrophoresis

by Anna Farkas

Supervisor: Dr. Guttman András

UNIVERSITY OF DEBRECEN
DOCTORAL SCHOOL OF MOLECULAR MEDICINE

DEBRECEN, 2023

Glycoanalytical identification of early differential diagnostic markers by capillary electrophoresis

by Anna Farkas, Molecular biologist MSc

Supervisor: Dr. András Guttman, PhD, DSc

Doctoral School of Molecular Medicine, University of Debrecen

Head of the **Defense Committee**:	János Szöllősi, PhD, DSc, Professor emeritus, member of the Hungarian Academy of Sciences
Reviewers:	Tamás Tábi, PhD
	Anett Türk-Mázló, PhD
Members of the Defense Committee:	Éva Szökő, PhD, DSc
	János Kádas, PhD

The PhD Defense takes place at the Lecture Hall of In Vitro Diagnostics Building, Faculty of Medicine, University of Debrecen, 16th of June, 2023, 11:00 a.m.

1. Introduction and objective

With the growing popularity of multi-site studies, it becomes necessary to standardize preanalytical conditions that are more difficult to control and standardize, and to assess the potential effects of these conditions on N-glycan content. As some of the samples I analyzed were not collected by various Institutions at the University of Debrecen, in the first part of the dissertation I wanted to address the key question whether the time elapsed from blood sample collection to centrifugation and freezing could influence the results of N-glycomic analysis of serum samples accomplished by capillary electrophoresis with lase-induced fluorescence detection (CE-LIF). The results may help to optimize the sampling protocols for the different studies in the future and to decide whether samples can be taken with sufficient safety for N-glycomic analysis in cases where a longer time elapses are between sampling and freezing (e.g. for infrastructural reasons) or when the time is a critical factor that can cause differences in the results of capillary electrophoresis measurements. These differences may appear as a source of errors in the evaluation of analysis.

Knowing the results of these preliminary experiments, in the further part of the thesis I deal with the investigation of the effect of obesity and gestational diabetes mellitus (GDM) during pregnancy on the maternal and child-derived serum IgG and IgA isolated from maternal serum, as well as the N-glycomic study of chronic obstructive pulmonary disease (COPD), lung cancer and the co-occurrence of the two lung diseases.

The choice of topic is justified by the fact that COPD, which is a common disease in Hungary as well, and will be the third leading cause of death worldwide by 2030, and lung cancer will take sixth place as forecasted by the World Health Organization (WHO) [1], in addition, obesity is a chronic condition that is also one of the most common health problems. In 2019, according to WHO data, obesity and its complications were the 9th leading cause of death globally. A common driver of the above listed diseases is the systemic chronic inflammation (SCI), as the study of the N-glycosylation pattern of IgG and IgA reflecting it, may be a valuable source of biomarkers.

Adipose tissue, as the largest endocrine organ, is able to secrete more than 50 adipokines, chemokines and cytokines, which affects the function of the immune system [2]. Inflammation associated with obesity and gestational diabetes during pregnancy can significantly affect the condition and transport processes in the placenta, and thus both the development of the fetus

and its health status after birth [3]. IgG contributes to the protection of the newborns already during fetal life in the womb and after birth, as long as its own antibodies appear in their circulation [4]. The N-glycosylation properties of the molecule affect its function, and its modifications may be markers for inflammatory processes [5], that may also occur during pregnancy associated with obesity and gestational diabetes. IgA antibodies, which are considered primarily attenuating of inflammatory responses, are absorbed through breastfeeding to support the proper functioning of the immune system. The study and identification of changes in the N-glycosylation pattern of IgG and IgA may be a means of investigating, effectively treating and preventing adverse effects in pregnancy, including offsprings.

In my research, I hypothesized that maternal obesity and gestational diabetes cause changes in the quality of placental transport processes and N-glycosylation of the two mentioned antibodies that can be detected on both maternal and offspring immunoglobulins, affecting health status of future generation at birth (birth weight, food sensitivity, level of C-reactive protein) and subsequent quality of life.

The high incidence rate of lung cancer in patients previously diagnosed with COPD suggests that the two diseases may have common pathomechanisms, such as premature aging in the lungs, genetic predisposition to the disease, activation of the common intracellular pathways, increased expression of growth factors, frequent development of respiratory diseases or common epigenetic factors [6]. For both respiratory diseases, the key factors to the effectiveness of therapy are the early diagnosis and early initiation of personalized therapy.

In the course of my PhD work, my aim was to develop a new glycomic method that would allow a simple, rapid and cost-effective study of changes in serum IgG and IgA N-glycosylation patterns. My aim was also to identify glycobiomarkers for the above-mentioned diseases by analyzing the N-glycan profile of serum samples, which could predict the development of COPD in different risk groups and allow the early detection of possible malignant transformation in patients with COPD.

Since the pathologies I examined typically show a close occurrence, my main goal during my work was to explore the glycobiomarkers of differential diagnostic significance of the diseases.

2. Materials and Methods

2.1. Reagents

The protein standards of IgA, haptoglobin (Hp), alpha-1-antytrypsin (A1AT) and transferrin (Tf) were from Molecular Innovations (Novi, MI, USA), while IgG standard human serum was purchased from Sigma Aldrich (St. Louis, MO, USA). Each of the standard glycoproteins was ≥ 95% pure according to the vendors. Sodium dodecyl sulfate (SDS) was from VWR (Radnor, PA, USA) and the RIPA Lysis Buffer from Merck (Darmstadt, Germany). The endoglycosidase PNGase F was from Asparia Glycomics (San Sebastián, Spain). The Fast Glycan Labeling and Analysis Kit was from SCIEX (Brea, CA, USA) including the 8-aminopyrene-1,3,6-trisulfonic acid (APTS) labeling dye and the magnetic beads for purification, maltooligosaccharide ladder and maltose as standard, HR-NCHO and NCHO separation buffers. Binding-specific exoglycosidases used to more accurately determine the glycan structures of the samples: Sialidase A (*Arthrobacter ureafaciens*), galactosidase (*Canavalia ensiformis*), and hexosaminidase were purchased from ProZyme (Hayward, CA, USA). Water (HPLC grade), acetonitrile, sodium cyanotrihydridoborate (1 M in tetrahydrofuran), glycerol, dithiothreitol (DTT), ammonium acetate (AmAc), sodium phosphate (NaH_2PO_4) and imidazole were from Sigma Aldrich (St. Louis, MO, USA). I purchased sodium chloride (NaCl) from MOLAR CHEMICAL Kft. (Halásztelek, Hungary). Protein G microaffinity columns and buffers used to purify IgG from serum samples and Ni-IMAC columns used to specifically capture IgA were ordered from PhyNexus Inc. (San Jose, CA, USA).

2.2. Peripheral blood samples collected for experiments for the investigation of the effect of storage at room temperature on N-glycosylation

To investigate the effect of storage at room temperature on the serum N-glycan profile, I applied the serum blood collection tube traditionally used in clinics, and the reference sample providing the N-glycan profile was treated according to the regulations. Venous blood samples from eight healthy young volunteers were collected into Vacutainer blood collection tubes containing coagulation activator (SiO_2) and a separator gel. Blood collection tubes were obtained from Becton, Dickinson and Company (Franklin Lakes, NJ, USA). Serum samples from 3 subjects were used to set the test parameters during the preliminary experiments, while after pooling

(mixing the same amount of samples), samples from all subjects included in the experiment (n=8) were again processed and analyzed using a capillary electrophoresis device. The control samples were kept at room temperature (24-25°C) for 30 minutes after blood collection and then centrifuged at 4°C for 20 minutes (2690 x g). This time is required for the coagulation process to happen, which was accelerated from an average of 60 minutes to 30 minutes in SiO_2-spray-coated blood collection tubes. All additional samples were kept at room temperature for 90; 150 and 270 minutes prior to centrifugation. The resulting serum fractions were aliquoted and stored at -20°C until further uses.

2.3. Preparation of serum samples for analysis of N-glycan profile change due to time-delay from sampling to centrifugation

Two µL of the serum samples obtained as described in 2.2 were diluted with HPLC grade water to 10 µL, and their glycoprotein content was denatured and reduced by adding 5 µL of denaturing solution (0.375% NP-40, 12.75% glycerol, 0.625% SDS and 12.5 mM DTT) applying temperature gradient denaturation procedure by increasing the temperature with the rate of 5.0°C/min from 30°C to 80°C [7]. The N-linked glycan structures of the proteins were released from their asparagine amino acid residues by enzymatic digestion with the addition of 1 µL of PNGase F (200 mU) and 19 µL of HPLC grade water during the individual sample preparation. 16.7 mM ammonium acetate (pH 6.9) buffer solution was used for the enzymatic release of N-glycans during the analysis of the pooled serum samples. The samples were incubated at 50°C for 1 hour followed by cooling down the mixture on ice for 1 minute, diluted with 70 µL of ice-cold acetonitrile in order to precipitate the protein content by centrifugation at 11.290 x g for 10 minutes. The released sugar containing supernatant was dried in a vacuum centrifuge (SpeedVac, Thermo Fisher Scientific, Schaumburg, IL, USA) and labeled using the fluorescent labeling solution shown in Section 3.2 in a heating block with open cap at 37°C overnight [8]. After the purification step with magnetic beads, the samples were handled as described in Section 3.2 of the dissertation.

2.4. Glycosylation analysis of IgG and IgA derived from maternal and child serum samples

The samples for glycosylation analysis were provided by the staff of the Albert Szent-Györgyi Clinical Center, Department of Obstetrics and Gynecology, Faculty of Medicine, University of Szeged, and the samples were collected in accordance with the ethical permit No. 172/2018-SZTE. Factors that may be associated with an increased risk of developing gestational diabetes mellitus (GDM) in addition to overweight /obesity and those that may affect the glycan profile have been identified as reasons for exclusion from the study. Since breast milk was collected in addition to serum and fatty acid analysis of the samples was performed with the help of the Hungarian University of Agricultural and Life Sciences, subjects were excluded from the study if:

1) the mother had type 1 diabetes mellitus
2) mother was on a special diet
3) a genetic disorder has been identified in the mother and / or unborn child
4) twin pregnancy
5) *Cesarean* delivery [9]

Blood samples were taken at 12 weeks postpartum. Maternal body mass index (BMI) and the presence of GDM were considered in the design of the study groups. Subjects with BMI ≥ 30 were in the obese group and BMI ≤ 30 in the normal weight group. Accordingly, participants were classified into the following groups:

1) Healthy Normal Weight (without GDM) – **HNW**
2) Healthy Obese (without GDM) – **HOB**
3) Normal Weight with GDM – **NWGDM**
4) Obese with GDM – **OGDM**

Prior to N-glycosylation analysis for each group, 15–15 samples were pooled by study group for both mother and offspring. Both maternal serum IgG and IgA and child serum IgG antibodies were purified by affinity chromatography.

2.5. Serum samples for IgG and IgA N-glycosylation analysis

Affinity microcolumns with a resin volume of 40 µL containing Protein G were used for specific purification of IgG antibodies and Ni-IMAC microcolumns (Nickel Immobilized Metal Affinity Chromatography) with a resin volume of 40 µL (PhyNexus Inc., San Jose, CA, USA) to capture the IgA from samples. Semi-automatic IgG and IgA purification steps were performed with PumpControl 2 software and an E4 XLS + electronic pipette, both provided by PhyNexus Inc. Protein purification steps were performed for IgG by optimizing the procedure published by Mesko et al. for the amount of sample available [10]. I used the procedure developed by Mészáros et al. [11] for purification of serum IgA after its modification. The Z(IgA1) affibody molecules containing 10 N-terminal His-Tags required for the specific binding of IgA were obtained from the Research Institute of Biomolecular and Chemical Engineering of the University of Pannonia.

2.5.1. Purification of IgG

Buffers:

1) 5 x Binding / Wash Buffer I (Buffer A): 50 mM NaH_2PO_4, 0.7 M NaCl (pH 7.4)
2) Wash buffer II. (Buffer P): 140 mM NaCl
3) Elution buffer: freshly prepared 10% acetic acid (pH 2.5)

Binding / Wash Buffer I (Buffer A) and Wash Buffer II (Buffer P) applied for the selective capture of IgG were part of the "Box of 96 PhyTip® columns (1 ml volume) containing 40 µl of Protein A affinity resin per column" distributed by PhyNexus Inc.

The samples were prepared as follows:

220 µL of 1x Buffer A was added to the 220 µL of serum samples then the mixture of samples and buffer was thoroughly vortexed. Prior to applying 400 µL of the mixture of sample and Buffer A per column those were conditioned using 200 µL of 1 x Buffer A. After the IgG binding step, the resin was washed with 200 µL of Buffer P followed by elution of the bound antibodies using freshly prepared 200 µL of 10% acetic acid. As the elution buffer was highly acidic, samples were transferred to a microcentrifuge tube containing a 10 kDa centrifugal filter for buffer exchange and volume reduction. Acetic acid was removed by centrifugation at 11,384

x g for 10 min. To remove any acetic acid remaining on the filter, 50 µL of HPLC grade water was added to the filter and centrifuged again at 11,384 x g for 10 min. Protein denaturation was performed (80°C; 10 min) on the filter by adding 10 µL of HPLC grade water and 4 µL of denaturing solution (400 mM DTT, 5% SDS). To avoid evaporation due to the high temperature in the thermostat, the tops of the microcentrifuge tubes were coated with paraffin in each case during the incubations. After removing the denaturing solution by centrifugation, the filter was washed with 30 µL of water by centrifugation at 11,384 x g for 10 min. The 10 kDa filters were transferred to clean microcentrifuge tubes and the N-glycan contents of the denatured and reduced proteins were released by the addition of 49 µL of 20 mM $NaHCO_3$ (pH 7.0) and 1 µl of PNGase F (200 mU) (37°C, overnight). The released glycans were washed away from the filter with 30 µL of HPLC grade water by centrifugation at 11,384 x g for 10 minutes. The carbohydrate-containing flow-through fractions were dried in a vacuum centrifuge (Thermo Scientific, Schaumburg, IL, USA). Lyophilized N-glycans were labeled with a mixture of 6 µL of 20 mM 8-aminopyrene-1,3,6-trisulfonic acid (APTS) (dissolved in 15% acetic acid) and 2 µL of 1 M sodium cyanoborohydride ($NaBH_3CN$) solution (1 M in tetrahydrofuran) by incubation at 37°C, overnight.

2.5.2. **Purification of IgA**

Buffers:

1) Z(IgA1) Binding buffer (Buffer A): 20 mM NaH_2PO_4, 50 mM NaCl (pH 7.4)
2) IgA Binding and Wash buffer (Buffer B): 50 mM imidazole, 20 mM NaH_2PO_4, 500 mM NaCl (pH 7.4)
3) Elution buffer (Buffer C): 20 mM NaH_2PO_4, 50 mM NaCl, 500 mM imidazole (pH 2.5)

900 µL Buffer B was added to 100 µL of samples. The columns were preconditioned with 400 µL of Buffer A and then functionalized with 1 mL 1 mg/mL His-tag Z(IgA1) affibody dissolved in Buffer A. After immobilization of affibodies to the columns, the resins were washed twice with 400 µL of Buffer B. After binding of IgA to Z(IgA1), the micro-columns were washed twice again with 500 µL of Buffer B to remove non-specifically bound molecules. The Z(IgA1) - IgA complexes were eluted from the columns using 200 µL of Buffer C. The samples were further handled according to the sample preparation steps described in Section 2.5.1.

2.6. Samples collected for the analysis of N-glycomic differentiation between COPD and lung cancer

Serum samples from 100 healthy individuals (control), 100 lung cancer patients (LC), 100 COPD patients and 100 patients with comorbidity of COPD with lung cancer were collected and pooled. The male/female ratios within the groups were as follows: 52%/48% in the COPD group, 64%/36% in the LC group, 72%/28% in LC+COPD group and 61%/39% in the healthy control group. The average age of the groups (mean ± SD) were 56.04 ± 14.46 in the control group, 64.42 ± 9.05 in the LC group, 66.20 ± 9.68 in the COPD group and 65.19 ± 8.38 in the comorbidity group, respectively. The patient and control serum samples were collected with the appropriate Ethical Permissions (approval number: 23580-1/2015/EKU (0180/15)) and Informed Patient Consents at the Pulmonology Department of Semmelweis Member Hospital of Borsod-Abaúj-Zemplén County Hospital and University Teaching Hospital (Miskolc, Hungary) and the Department of Laboratory Medicine, University of Debrecen (Debrecen, Hungary). Patients who had 1) any other pulmonary disease, 2) lung cancer as a metastasis, 3) autoimmune diseases, 4) drug-addiction were all excluded from the study.

2.7. Preparation of serum and standard glycoproteins for N-glycosylation analysis

For the individual measurements of IgG, IgA, haptoglobin, alpha-1-antitrypsin and transferrin, 10 μL of 10 mg/mL protein standards were used.

A standard protein mixture (SPM) was prepared to model the N-glycosylation profile of human serum. The concentration of the individual stock solutions for the standard protein mixture was 10 mg/mL except for IgG, which was used in 50 mg/mL concentration. The SPM contained the 5 glycoproteins mentioned above in the following target concentrations:

1) Haptoglobin (Hp) – 2.0 mg/mL (0.3-2.0 mg/mL)
2) Transzferrin (Tf) – 3.0 mg/mL (2.0-3.6 mg/mL)
3) Alfa-1-antitripszin (A1AT) – 1.0 mg/mL (0.9-2.0 mg/mL)
4) IgA -2.1 mg/mL (0.7-4.0 mg/mL)
5) IgG – 7 mg/mL (7-16 mg/mL)

The physiological serum concentrations of the proteins are shown in brackets after the target concentration [12].

The final concentration of the individual glycoproteins in the standard protein mixture (SPM) was fine-tuned experimentally until the N-glycome profile became similar to that of the human serum. The purity of the standards was evaluated by sodium dodecyl sulfate polyacrylamide slab gel electrophoresis (SDS-PAGE) using precast Tris-Glycine 10-20% polyacrylamide gradient gels (Thermo Fisher Scientific, Carlsbad, USA) following the manufacturer's protocol.

Two microliters of serum samples were diluted with HPLC grade water to 10 µL, followed by denaturation at 65°C for 10 min using 5 µL denaturation solution (60 µL 1x RIPA buffer (0.5 M Tris-HCl pH 7.4, 1.5 M NaCl, 2.5% deoxycholic acid, 10% NP-40 and 10 mM EDTA), 10 µL of 5.0% SDS and 10 µL of 100 mM DTT). Release of the N-glycans from the denatured proteins was performed by the addition of 1.0 µL of PNGase F (200 mU) and 50 µL of HPLC grade water. Temperature and pH-sensitive α(2-3,6,8,9)-linked sialic acid structures of the N-glycans were removed by addition of 1.0 µL of *Arthrobacter ureafaciens* sialidase A enzyme (1 U) and the mixture was incubated at 50°C for 60 minutes. After the sugar release step, the samples were cooled on ice for 1 minute and diluted with 120 µL of ice-cold acetonitrile in order to precipitate the protein content of the samples, followed by centrifugation at 11,290 x g for 5 minutes. The released glycans containing supernatant was dried in a vacuum centrifuge. 4 µL of 40 mM APTS (in 20% acetic acid), 2 µL of $NaBH_3CN$ (1 M in THF) and 4 µL of 20% acetic acid were added to the dry samples and the reaction mixture was labeled using evaporative fluorophore labeling procedure in a heating block with a closed cap at 50°C for 60 min and with an open cap at 55°C for 60 min [8]. After the labeling step, the samples were purified by magnetic beads from the excess fluorophore following the Fast Glycan Sample Preparation and Analysis Protocol (SCIEX). 87.5% acetonitrile was added to the magnetic beads and labelled sugars contained mixture during this purification step. This concentration is ideal concentration for the binding of sugars to magnetic beads [13]. The labelled glycans were eluted from the beads with HPLC grade water and analyzed by capillary electrophoresis with laser-induced fluorescence detection or stored at -20°C.

2.8. Capillary electrophoresis

Capillary electrophoretic analyses were performed using a PA800 Plus Pharmaceutical Analysis System (Beckman Coulter Inc., Brea, CA, USA) with a laser-induced fluorescence detector. To detect the N-glycan content of the samples, the APTS tagging dye was excited with light at a wavelength of 488 nm provided by an Ar-ion laser, and the light emitted by the dye was collected with an emission filter (520 nm). Bare fused silica (BFS) capillary with an effective length of 50 cm (60 cm full length) and an internal diameter of 50 μm was used during all N-glycan separations. The capillaries were filled with HR-NCHO (pH 4.75) or NCHO (pH 4.75) separation gel buffers before sample injection, into which the sample was injected electrokinetically in case of HR-NCHO gel and pressure in case of the NCHO gel. The equipment was used in reverse polarity mode (cathode at the injection side, anode at the detection side) during separations. The separation conditions applied at the time of analyses were indicated in each case under the electropherograms.

2.9. Data evaluation

32Karat (version 10.1) software package (Beckman Coulter Inc., Brea, CA, USA) was applied for data acquisition and processing. The peaks of the electropherograms obtained as a result of the separations can be assigned to N-glycan structures were based on their glucose unit (GU) values. The GU values were calculated with the GUcal software (www.gucal.hu) [14]. N-glycan structures were assigned to the peaks using the built-in database of the software, based on literature data and by searching in an external database. Peaks of the electropherograms with the average relative area under the peak reached the minimum value of 1% (area percentage ≥1%) were considered as important during the identification of the N-glycan content of the examined samples. Normalized peak area percent values of the separated components were calculated using PeakFit v4.12 software (SeaSolve Software Inc., San Jose, CA, USA).

2.10. Statistical analysis

Mean values along with standard errors of the data obtained were determined using PeakFit software at the time of analysis. Data are presented as mean ± SD in the text. Statistical analysis was performed with GraphPad Prism (GraphPad Software, San Diego, CA, USA; version 8.0.1). Shapiro-Wilk test was applied to examine the distribution of the data in all cases.

Friedman test [15] supplemented with Dunn's post hoc test was used to investigate the effect of time at room temperature on the N-glycan composition of the sample, as not all peaks showed a normal distribution [16]. The evolution of the sialic acid/neutral ratio of the N-glycan structures of the samples was examined using Friedman's test and Dunn's post hoc analysis. In the case of the pooled samples, due to the increase in standard deviation observed when analyzing the ratio of sialic acid/neutral structures in the sample stored for 150 minutes, I performed a ROUT outlier analysis (Q=1%).

Kruskal-Wallis test followed by Dunn’s test was applied to examine the differences between the groups during the N-glycosylation analysis of maternal and child serum IgG and IgA immunoglobulins. [17]. The sialic acid / neutral ratio of the N-glycan structures of the samples was examined by one-way analysis of variance (One-way ANOVA) and Tukey's range test for the normal distribution of the data, otherwise by the Dunn's test.

Kruskal-Wallis test was performed for statistical analysis of results obtained from samples derived from control, patients with COPD, lung cancer, and co-occurrence of the two diseases, whereas the normal distribution could not be determined with certainty due to the reduced number of samples resulting from their pooling. Dunn's test as a *post hoc* analysis was used to determine the differences between the study groups Differences between the groups were considered as significant at $p \leq 0.05$, which were denoted by * in the figures as follows: * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$; **** $p \leq 0.0001$.

3. Results

3.1. Changes in human serum N-glycan profile due to storage at room temperature (24-25°C)

13 major N-linked glycan structures with at least 1% area percent were identified during the individual (n = 3) glycosylation analysis of serum samples. It was found that storage at room temperature had a minimal effect on the total N-glycan profile of the serum. Since the focus of my interest was how sialic acid-terminated structures change with increasing time until centrifugation and freezing, I defined the ratio of the (SF – sialoform) to the neutral structures (NF – neutral form) (SF/NF ratio). From the point of view of the development of the ratio of sialic acid structures (SF/NF ratio), I found that the proportion of sialylated structures showed a slight but continuously decreasing trend over time, which slightly increased again in case of the 270-minutes sample. However, subjecting the development of SF/NF ratios to statistical analysis, I did not find any significant differences between the groups.

In order to take into account possible individual differences, I studied the the glycan profile obtained by pooling the serum samples of 8 healthy volunteers. Ammonium acetate (16.7 mM, pH 6.9) was used as buffer for the enzymatic release of N-glycans from proteins, and glycan analysis was performed using HR-NCHO separation buffer providing a higher resolution. 16 glycan structures were identified during the evaluation of the measurement data. During the examination of the SF/NF ratio, a decrease in the sialic acid content of the 90- (3.35 ± 0.46) and 270-minutes (3.07 ± 0.18) samples was observed comparing to the control in the case of the pooled samples, however, serum samples stored for 150 minutes (7.48 ± 2.02) showed an increase in SF/NF ratio that exceeded the value of the control group (3.73 ± 0.33). However, the rate of increase was not statistically significant compared to the control ($p=0.1708$), in contrast to the 90-minutes ($p=0.0115$) and 270-minutes ($p<0.0001$) samples. However, the SF/NF ratio showed no significant difference between the control and 90-minutes ($p>0.9999$), control and 270-minutes ($p=0.1057$) and 90- and 270-minutes ($p=0.8648$) samples. Statistical analysis was performed using Tukey's test supplemented by Dunn's *post hoc* analysis.

During the statistical analysis of the N-glycan structures of the pooled samples by Friedman and Dunn tests, I found a statistically significant difference in 6 of the 16 identified peaks, however, none of these differences occurred between the control time point and any other time

point examined. All significant differences were observed between time points other than the control.

3.2. Effects of maternal obesity and gestational diabetes

3.2.1. Comparison of N-glycosylation of IgG isolated from maternal and child serums

In the second stage of my PhD work, I studied the N-glycosylation modifications of IgG antibodies purified from maternal and child serums using Protein G affinity chromatography columns.

Examining the N-glycosylation of IgG antibodies isolated from a group of healthy normal weight (control) mothers, I was able to identify 14 structures that met the criteria I previously defined and described above (mean relative area percentage ≥ 1%). Examining the IgG-derived glycan profile of serum samples from children, 13 peaks met the same condition.

When comparing the ratio of sialic acid-neutral structures by Kruskal-Wallis test and Dunn's post hoc analysis, I found a statistically significant difference between healthy, normal-weight mothers and their children (HNW mothers SF/NF=0.33 ± 0.01; children of HNW mothers SF/NF=0.30 ± 0.03; p=0.026) and healthy but overweight mothers and their children (HOB mothers SF/NF=0.29 ± 0.02; children of HOB mothers SF/NF=0.34 ± 0.02; p=0.0087).

3.2.2. Effect of pregnancy complicated by maternal obesity and gestational diabetes mellitus on N-glycosylation of IgG isolated from maternal serums

During the statistical analysis of the SF / NF ratio with the Kruskal-Wallis test, I found only a small but significant difference between the HOB and NWGDM groups (p=0.0197). The HNW group showed no significant differences between the HOB (p = 0.2699), NWGDM (p> 0.9999), and OGDM (p> 0.9999) groups. In addition, there was no significant difference between OGDM and HOB (p = 0.2727) or OGDM and NWGDM (p> 0.9999). In addition to examining the proportion of sialic acid-terminated structures, the relative area percentage value for each peak were subjected to statistical analysis. Data were analyzed using Kruskal-Wallis test followed by Dunn's post hoc analysis, as a result of which all structures except for 3 of the 14 maternal N-glycan structures identified (peaks 5, 7, and 11) showed significant differences

between groups. Of the 11 structures that showed significant differences between any of the groups studied, 10 structures differed between the control and any groups. Of the 10 structures, the control and HOB groups can be distinguished by 4 structures (A2G2S2, p=0.0003; FA2G2S2, p=0.0006; FA2BG2S2, p=0.0238; FA2 [6]G1, p=0.0133), the control and NWGDM groups by 2 structures (FA2B, p=0.0003; FA2B [6]G1, p=0.0098), and the control and OGDM groups by 7 structures (A2G2S2, p=0.0314; FA2G2S2, p=0.0291; FA2BG2S2, p=0.014; FA2(3)G1S1, p=0.0193; FA2G2S1, p=0.0088; FA2G2, p=0.0002; FA2BG2, p=0.003).

3.2.3. Effect of maternal obesity and gestational diabetes mellitus on N-glycosylation of IgG from serum samples of offsprings

In the N-glycosylation study of IgG isolated from the blood of children born to the mothers of the study groups, I identified 13 peaks with a relative area percentage of at least 1%. The SF/NF ratio was significantly higher in the group of children born to HOB mothers compared to the children of control (p=0.0423) and children of OGDM mothers (p=0.0225). I also found a significant difference between the children of the maternal groups diagnosed with the presence of GMD. The OGDM children were characterized by a significantly lower degree of sialylation than children of normal weight mothers with gestational diabetes (p=0.033). There was no significant difference between the HNW and NWGD (p=0.0607), HNW and OGDM (p>0.9999) and between the HOB and NWGDM (p>0.9999) groups. Among the 13 identified N-glycan structures, the relative area percentages of 9 structures showed statistically significant differences between groups. Of these 9 structures, I was able to differentiate the control group from the HOB group based on 3 structures (FA2(3)G1S1, p=0.0076; FA2[6]G1, p=0.0151; FA2B[6]G1, p=0.033) and from the NWGDM group based on 4 structures (FA2G2S2, p=0.0115; A2G2S1, p=0.0225; FA2[6]G1, p=0.0003; FA2[3]G1, p=0.0197).

3.2.4. Effect of maternal obesity and gestational diabetes mellitus on N-glycosylation of maternal serum IgA

During the N-glycomic analysis of IgA antibodies isolated from maternal serum was identified 19 peaks with a relative area percentage of at least 1%. I assigned the 20 N-glycan structures (structures 9 co-migrated with structure 10) to these peaks based on their GU values, database search and literature data. A significant proportion of the identified N-glycan structures, 15 of

the 20 structures, showed statistically significant differences between at least two groups compared. However, contrary to expectations, IgA glycosylation of the obese group could only be distinguished from the control group by a single N-glycan structure (A2 [3] G1S1, p=0.0042), in contrast to IgG antibodies, where 4 structures showed significant differences between the control and obese groups. During comparison of the control group and group of normal body weight diagnosed with gestational diabetes, 3 structures differed significantly (FA1[6]+A2BG2S1, p=0.0291; A1G1[6], p=0.0423), while 7 structures differed significantly (FA2BG2S2, p=0.0291; A[3], p=0.0132; M5, p=0.0027; A1G1[6], p=0.0006; FA1[3]G1, p=0.0225; A2G2, p=0.0197 and FA2G2, p=0.0065) when comparing the control group with the group of obese with gestational diabetes. It should be noted, that a significant proportion of the peaks showing a significant difference between the control and the group of obese gestational diabetes mothers were peaks with a smaller area percentage, with an average value not exceeding 4%. Therefore, the ODGM group can be safely separated from the control group on the basis of area % of FA2BG2S2, A [3], and A2G1 [6] structures, of which the monogalactosylated structure was emergent. However, I did not find any significant differences between the groups in terms of the trend of sialylation structures indicating inflammatory processes.

3.3. N-glycomic analysis of COPD and lung cancer using human serum derived from whole blood

3.3.1. Comparison of standard human serum and a mixture of standard proteins present in high concentrations in the serum

Different tissue and cell types may express different protein glycoforms under physiological and pathological conditions that may have diagnostic significance after entry into the circulation. In the field of tumor diagnostics, such glycoproteins may appear to be a very attractive biomarker target due to their specificity, although these proteins are expected to be present in the bloodstream pg / mL concentration range [18]. Therefore, great emphasis should be placed on the development of sufficiently sensitive methods that can greatly increase the time, material, and sample requirements. Although depletion of high abundant serum proteins is more likely to reveal potential biomarkers at low concentrations, several studies reported that

known serum components and acute phase proteins may also carry diagnostically relevant information [19], not to mention the analysis of whole serum/plasma samples [20]. In addition, the specific purification of various proteins from serum for glycosylation analysis can be accomplished by time-consuming and costly procedures that require a large sample amount compared to direct analysis of serum. To eliminate this, I have developed a new potential method to obtain sufficient information on the N-glycosylation of the two antibodies present in the highest concentration in human serum and three additional glycoproteins by examining the serum sample as a whole. During N-glycomic analysis of commercially available standard human serum after desialylation and desialylated model serum (Standard Protein Mixture - SPM), which was constructed by mixing the five high abundant human serum glycoproteins at their respective physiological concentrations, the same nine major (mean value of the relative peak area ≥1%) desialylated N-linked glycan structures were identified. Apparently, peaks 1-4 (FA2, FA2B, FA2(6)G1, FA2(3)G1) originated from IgG, peaks 5 (A2G2) and 6 (FA2G2) from each of the five high abundant proteins. Peak 7 (FA2BG2) is derived from both IgG and IgA, whereas the source of the three-antennary structures, i.e., peaks 8 (A3G3) and 9 (FA3G3), is presumably the haptoglobin protein. It must be noted that the five proteins chosen for the SPM roughly corresponded to ca. 85% of the glycosylated proteins of the human serum, therefore, as a first approximation, I considered that the minor differences were caused by the remaining ~ 15% of lower abundant serum glycoproteins [21].

3.3.2. N-glycosylation investigation and statistical analysis of lung cancer, COPD and comorbid patient groups

In the study of patient groups, after the desialylation step, which increased the stability of the serum N-glycan profile, similar to the standard serum and the SPM sample I was able to identify 9 major glycan structures based on GU values of the peaks. During the detailed examination of the glycan structures of serum glycoproteins an increase in the degree of branching (Branching Degree – BD) was previously observed under pathological conditions compared to the control [22]. Since the degree of branching can be of diagnostic importance even without sialic acids, I determined the degree of branching in the samples of control, COPD, lung cancer, and comorbid groups using the relative percentage distribution of tri-antennary glycan structures in the serum samples as follows:

Branching Degree = Tri-antennary glycans (%) / Total glycans (%)

During the evaluation of the degree of branching, I observed an increase in all examined patient groups compared to the control group. Among COPD patients, the ratio of tri-antennary structures (0.07) showed a notable but not statistically significant increase compared to the control group (0.04; p=0.2492) and comorbid group (0.05; p>0.9999). The difference between COPD and lung cancer (0.09) was also not significant (p>0.9999). Although the value of each group increased compared to the control, the most remarkable increase was observed for the lung cancer (0.09) group, where the difference between the two groups was also statistically significant (p=0.0134). The increase in branching rate was mainly due to a small increase in peak area % value of the FA3G3 structure in the comorbid group, while this increase was caused by a significant increase in the same glycan structure and different changes in all mono- and bigalactosylated structures in the COPD and lung cancer groups.

I also determined the relative change in the percentage of the area of the peaks identified in the examined patient groups in value of % compared to the same structures of the control group. During the analysis, an opposite alteration was observed in the relative peak area % of the FA2 structure, which can thus be a potential molecular marker to effectively differentiate COPD (-37.47% ± 1.12) from lung cancer (+12.33% ± 0.22) and both lung diseases can be distinguished from the healthy control group by examining the structure. The relative peak area % of comorbid group decreased only by a very small extent for the same FA2 peak compared to the control group (-0.60% ± 0.01). Negative changes in relative peak area % of the bi-antennary core-fucosylated bisecting FA2B and the core-fucosylated bi-antennary monogalactosylated FA2 [6] G1 and FA2 [3] G1 was observed in COPD (FA2B: -49.18 ± 1.34 ; FA2 [6] G1: -49.09 ± 0.63; and FA2 [3] G1: -44.92 ± 1.20), lung cancer (FA2B: -36.19 ± 1.30; FA2 [6] G1: -22.49 ± 0.26; FA2 [3] G1: -19.84 ± 1.52) and in the comorbid groups (FA2B: -28.22 ± 1.05; FA2 [6] G1: -34.21 ± 0.03; FA2 [3] G1: -30.57 ± 1.10) compared to the control group. A small negative change was observed in all three groups compared to the control (COPD: -15.94 ± 0.58; lung cancer: -23.68 ± 0.79; comorbid: -20.04 ± 0.56) in the case of the FA2G2 structure. Disease groups can be distinguished from the control group by examining these 4 glycan structures. It is important to note that based on SPM model and serum analysis, peaks FA2, FA2B, FA2[6]G1 and FA2[3]G1 were probably originated from IgG. The largest relative increase in peak area (+140.34% ± 3.73) was observed in the core-fucosylated tri-antennary trigalactosylated FA3G3 structure for the group of patients with lung cancer compared to the healthy control. The same structure differed by +81.70% ± 0.69 for COPD and only by +22.70%

± 1.53 for comorbidity patients. A notable increase was observed in the relative peak area percentages in the A3G3 peak of the COPD (+41.28% ± 1.19) and lung cancer groups (+28.44% ± 0.82), while those were slightly decreased for the comorbidity patients (-1.83% ± 0.14). These two core-fucosylated tri-antennary structures were presumable originated from the acute-phase protein of haptoglobin based on SPM and serum analysis. Although the A2G2 structure did not show significant differences between the groups in the subsequent statistical analysis, it was positively changed in all disease groups compared to the control during the first data analysis procedure. A2G2 and FA2G2 structures were apparently derived from all five proteins in the SPM model and, consequently contributed a very large proportion to the area % of these peaks in serum samples as well. The most stable glycan structure was considered to be FA2BG2 according to this method of analysis, which changed by less than a 10% in all diseases groups compared to the control (COPD: +7.02% ± 0.45; lung cancer: -2.66% ± 0 .03 and comorbid: +1.21% ± 0.04). Based on the available data, I considered this peak may be derived from IgG and IgA immunoglobulins. During the statistical analysis of the area percent values of the N-glycan structure of the serum samples, I found a significant difference in the mean area percentage of 3 N-glycan structures for COPD group (FA2B, p= 0.0134; FA2[6]G1, p=0.0134; and FA2[3]G1, p=0.0134)) and 2 structures for lung cancer (FA2G2, p=0.0194; and FA3G3, p=0.0134) compared to the control group. The two disease groups were also differentiated by 2 (FA2, p=0.0134 and FA2BG2, p=0.0194), and the COPD and comorbid groups by only one N-glycan structure (A3G3, p=0.0134). A2G2 N-glycan structure was the only neutral sugar structure detected in human serum that did not show significant differences between groups.

4. Discussion and Conclusions

4.1. Effect of storage at room temperature on the total N-glycan profile of human serum

Regarding the individual variability of N-glycome, the time and condition of sample collection (e.g., storage temperature and time) and the factors involved in sample preparation may be of great importance, so the central question in my experiments was whether storage of whole blood at room temperature affects the global N-glycan profile of serum samples obtained after centrifugation. In the first part of the work, the change of the total N-glycan content of 3 serum

samples collected from healthy volunteers caused by room temperature was investigated by applying capillary electrophoresis, during which I identified 13 major N-linked glycan structures. The ratio of sialic acid-terminated structures, as expected, showed a small but continuously decreasing trend with increasing storage time at room temperature (30 min, 90 min, 150 and 270 min) up to the 150 min time point. However, for the sample of 270-minutes, the SF/NF ratio increased and approached the initial value belonging to the control sample (30 min).

Studying the glycan profile obtained by pooling the samples of serum 8 healthy volunteers, I was able to identify 16 asparagine-linked glycan structures in the pooled samples with ≥ 1% relative peak areas. As the time spent at room temperature increased, the SF/NF ratio also showed a decreasing trend in pooled samples. In the case of the serum stored for 150 minutes, the SF/NF ratio exceeded the value of the control sample. The increase was significant compared to the samples of 90 min and 270 min.

During the statistical analysis of the samples, I found a statistically significant difference in 6 of the 16 identified peaks, however, none of these differences occurred between the control and any of the other time points examined. The significant changes were found at different time points, other than control in all cases.

Conclusion: A dynamic change in SF / NF ratio was observed as a function of time delay from blood collection to centrifugation and freezing of serum samples. Several factors can explain this phenomenon:

1) free sialyltransferases get activated in the serum
2) glycan structures released from the surface of lymphocytes and erythrocytes cause a shift in the SF / NF ratio
3) N-glycans released from intracellular organelles due to possible hemolysis appear in the serum over time, which is later degraded by free exoglycosidases [23].

All of these processes may be responsible for the changes observed in the ratio of sialic acid-neutral structures over time. Therefore, in order to examine the glycan profile free from pre-analytical factors, I considered to centrifuge the samples as soon as possible (following the regulations), to separate the serum from the other blood components and to store it frozen until downstream N-glycosylation analysis. If this does not happen, the results obtained do not always reflect the current N-glycosylation status at the time of blood collection.

However, my results suggested that storing the whole blood sample at room temperature until centrifugation had no significant effect on serum N-glycome beyond increasing the standard deviation within the study groups, if the simulated sample handling error (prolonged storage at room temperature before centrifugation) did not occur in a notable proportion of the samples collected. In order to control the increase in standard deviation caused by the conditions examined in the dissertation, pooling of samples may provide a solution if the individual analysis and comparison of patient data is not our goal. In addition to reducing the standard deviation, pooling significantly reduces the time and cost requirements of sample processing, while at the same time highlighting the N-glycosylation properties, which are most characteristic of a given group. As a result of my work, it can be stated that serum N-glycome can be considered as a relatively stable source of biomarkers under the conditions studied.

4.2. Effect of pregnancy complicated with obesity and gestational diabetes on N-glycosylation of IgG isolated from maternal and child serum samples

4.2.1. N-glycosylation pattern of IgG isolated from maternal and child serums

As a first step in my work, I determined the N-glycan structures of IgG antibodies in maternal and child serum samples using CE-LIF. I identified 14 N-linked sugar structures on IgG purified from maternal and 13 infant serum samples. During comparison of the maternal and child SF/NF ratio, I found statistically significant differences between healthy control mothers and their children (HNW, $p=0.026$) as well as obese mothers and their children (HOB, $p=0.0087$).

Conclusion: Similar to the control, I found a significant difference in the SF/NF ratio of N-glycosylation of IgG isolated from obese maternal and child serum, which ratio change is opposite to the control group (SF/NF ratio of HNW mothers $= 0.33 \pm 0.01$ and SF/NF ratio of their children $= 0.30 \pm 0.03$; SF/NF ratio of HOB mothers $= 0.29 \pm 0.02$ and SF/NF ratio of their children $= 0.34 \pm 0.02$), so I hypothesized that the transfer of maternal IgG antibodies across the placenta was impacted by maternal obesity and GDM. The results supported the hypothesis that the placenta may be able to “sense” the maternal environment and adapt to them

in order to protect the fetus from harmful inflammatory, oxidative, hyperinsulinemic environment present in maternal obesity and gestational diabetes [24].

Thus, taking into account the results of previous studies, I believe that a subclass and glycosylation site-specific IgG assay should be performed in the future with additional subjects, supplemented by an analysis of placental inflammatory markers. These studies can also lead to a better understanding of the transport processes of antibodies through the placenta and thus the factors influencing the development of the fetal immune system.

4.2.2. Effect of maternal obesity and gestational diabetes on N-glycosylation of maternal serum IgG

During the glycomic analysis of IgG isolated from maternal serum, significant differences were found in 11 of the 14 structures identified between the groups. Of these, 10 structures differed significantly between the control and a mother group, complicated by obesity, GDM, or both.

Conclusion: When analyzing the SF/NF ratio, although I observed a significant but small decrease only in the sialylation of the HOB group compared to the values of NWGDM group ($p=0.0197$), I also found a decrease in the SF/NF ratio of the OGDM group compared to the control and NWGDM groups. The results, according to the ratio in the group of mothers with gestational diabetes did not change at all compared to the control, and the fact that the difference was statistically significant only between the NWGDM and HOB groups showing that obesity rather than gestational diabetes had an effect on the N-glycosylation of IgG in mothers. Thus, the resulting SF/NF ratio may indicate inflammatory processes occurring in obesity, but at the same time, the possible reason for the small changes may be the gradual return of IgG glycosylation to the pre-pregnancy state after delivery [25], due to which the possible changes caused by GDM were less noticeable in the two affected group. However, the results obtained by the investigation of the GDM group were consistent with the results of a recently published research, according to which it was not possible to separate the groups of normal and GDM mothers considering changes in IgG N-glycosylation [26], therefore, I can confirm based on my results that the glycosylation analysis of IgG is probably not suitable for predicting the development of GDM. In the case of rheumatoid arthritis, which was also accompanied by chronic inflammation, in parallel with the improvement of the disease state during pregnancy, an increase was reported in galactosylation and sialylation of N-glycan structures linked to the

conserved glycosylation site in IgG Fc region [27] [28], thus, it can be assumed that the harmful effects of obesity and GDM were partially compensated by the hormonal changes occurring in pregnancy that could possibly be revealed with IgG N-glycosylation studies.
However, since the OGDM group had a higher SF/NF ratio than the HOB group and I observed a notable difference in the N-glycosylation of IgG that passed through the placenta from mother to the children, based on the results, I consider to be justified to implement further studies with larger groups of mother-child pairs. Since the FABG2S2 structure showed a significant difference between the control and HOB, as well as the control and OGDM groups, the structure may reflect the effects of obesity, and its follow-up is recommended during possible further studies.

4.2.3. Effect of maternal obesity and gestational diabetes on N-glycosylation of IgG antibodies isolated from serum samples of children

In the case of children born to the mothers of the studied groups, I found a statistically significant difference between the groups in the relative area percentages of 9 out of 13 identified complex N-linked sugar structures. I was able to separate the control group from the groups of children born to HOB mothers based on are percent of 3 structures (FA2(3)G1S1, FA2[6]G1, FA2B[6]G1), and 4 N-glycan structures differed significantly in the control and the groups of children born to NWGDM mothers (FA2G2S2, A2G2S1, FA2[6]G1, FA2[3]G1). A slight increase was experienced in the SF/NF ratio of both children of obese mothers and the children born to GDM mothers with normal body weight compared to the value of the control group, while I received the same value in the comorbid group showing both disorders as in the control group. There was a significant difference between HNW and HOB ($p=0.0423$), HOB and OGDM ($p=0.0225$), and NWGDM and OGDM ($p=0.033$) groups.

Conclusion:

In the light of the data, it can be concluded that:

1) Both obesity and GDM can influence the sialylation of N-glycosylation of IgG molecules detected in children.
2) It can be assumed that the transport of antibodies with a more anti-inflammatory effect was preferred by the placenta, compensating for the potentially inflammatory environment of the uterus, since IgG N-glycans of children born from pregnancies with obesity and GDM have an increased SF/NF ratio. The increase in the proportion of sialic acid-terminated structures mediates the anti-inflammatory effects of IgG.
3) Glycosylation-selective IgG transport, which can serve to compensate for the inflammatory environment of the uterus, may be part of this defense mechanism [29]. As a result of this process, the placenta may favor molecules with higher sialylation under inflammatory conditions.
4) The reduced sialylation observed in the children of the OGDM group may indicate damage to this mechanism.

The results urge for further studies to understand the effects of diseases occurring during pregnancy on placental transport processes and their effects on the offspring.

4.3. Changes in IgA N-glycosylation due to obesity and gestational diabetes

Since Steffen et al. previously showed that there was a correlation between the N-glycosylation profile of serum IgA and the pro-inflammatory effect of the molecule [30], the glycosylation modifications of IgA, similarly to IgG, can serve as an effective marker for various inflammatory processes occurring in the body during obesity and GDM. During my study, 15 complex sugar structures of IgA N-glycans showed statistically significant differences between at least two groups. The group of obese mothers could be differentiated from the control group on the basis of a single N-glycan structure (A2[3]G1S1, p=0.0042). The control and NWGDM groups could be distinguished according to 3 structures (FA1[6]+A2BG2S1; A1G1[6]), while the control and OGDM groups of mothers could be distinguished from each other according to 7 structures (FA2BG2S2; A[3]; M5; A1G1[6]; FA1[3]G1; A2G2 and FA2G2). Regarding the SF/NF ratio, I did not observe any significant differences between the groups.

Conclusion:

Based on my results, the following conclusions can be made:

1) Since significant differences were observed in the relative area percent of A1G1[6] structure in NWGDM compared to the control, as well as between the control and OGDM groups, this structure may be applicable for predicting the development of GDM regardless of BMI.
2) The A2G2, FA2G2, and M5 structures may be suitable for predicting GDM associated with obesity, since their percentage values differed significantly only between the HNW and OGDM groups. I recommend further investigation of these structures.
3) I concluded that the entire N-glycosylation profile of IgA is less affected by obesity than GDM, and accordingly, it can be a potential marker for the GDM development and its possible permanent effects, regardless of the mother's BMI.
4) The sialylation of IgA antibodies was not affected by either obesity or GDM, so the molecule was not useful for indicating the existence of inflammatory processes in the examined cases.

4.4. Modeling of serum N-glycosylation profile and application of the new approach in N-glycosylation analysis of various lung diseases

In order to significantly reduce the amount of sample required for the N-glycosylation analysis of IgG and IgA immunoglobulins, as well as the analysis time, I developed a new potential method, by which one can obtain sufficient information about the glycosylation modifications of these two antibodies and the most abundant serum glycoproteins by examining the entire serum sample. Using the SPM model, I successfully modeled the N-glycan profile of human serum treated with sialidase. Through the developed method (SPM), it is possible to study the glycosylation property of three more high-concentration glycoproteins of the serum in addition to the two immunoglobulins.

After the creation of the model, I studied the changes in the serum N-glycome of healthy controls, patients with COPD, lung cancer, and patients suffering from both diseases in order to effectively separate the diseases from each other. Similarly to the standard human serum used

as a control for the development of the method and to the model, I identified 9 N-glycan structures in the patients' serum samples after digestion of the whole profile with sialidase enzyme. During the evaluation of the results, I determined the degree of branching of the total serum N-glycome, during which I found a statistically significant difference only between the control group and the group of patients suffering from lung cancer ($p=0.0134$). In order to better illustrate and make it easier to understand the changes experienced between the individual peaks of the groups, I determined the degree of relative changes compared to the same structures of the control group. My results revealed that the FA2 structure can be a potential glycobiomarker that can be used to effectively differentiate COPD from lung cancer, and both lung diseases from the healthy control group. During the statistical analysis of the results, 3 structure differed significantly between the control and COPD groups (FA2B, FA2[6]G1 and FA2[3]G1), while the average area percentage of two N-glycan structures showed significant difference in the control compared to the lung cancer group (FA2G2 and FA3G3). The two disease groups (COPD and lung cancer) could be differentiated from each other using two glycan structures (FA2 and FA2BG2), and the COPD and comorbid groups could only be separated based on a single N-glycan structure (A3G3).

Conclusion: Based on the results of my experiments with the standard protein mixture, the glycosylation changes of the 5 high abundant glycoproteins present in human serum can be studied with a good approximation by analyzing the total human serum, thereby accelerated the analysis time and significantly reducing the sample and cost requirements of the tests. Accordingly, the information content relayed by the total serum glycan profile was sufficiently informative and there was no need in many cases to isolate and analyze individual proteins separately. The individual differences caused by the presence of the remaining approximately 15% were negligible if the protein to be studied was present in a sufficiently high concentration in the serum, i.e. if we assumed the presence of glycosylation changes in one of the 5 proteins also examined in the thesis. At the same time, even in this case, it may be worth to study the glycosylation of the protein even after specific purification, if differences are observed in the whole glycan profile of the serum compared to the control. Certainly in many cases we can reveal differences indicating to pathological processes even without procedures which is complicated, time consuming, costly and require experience/skills. In all cases, this is subject to a prior specialist opinion and knowledge of the suspected disease (we must know the examined patient presumably suffers from some kind of lung disease, in this case COPD). Accordingly, during the regular tests for patient care, it may be worthwhile to monitor the total

N-glycan profile of the serum in order to be able to identify a possible abnormal cell proliferation among the risk groups at an early enough stage, even before it can be detected by imaging methods. This may also be crucial in terms of treatment and the patient's chances of survival. Unfortunately, there is no regular and organized lung cancer screening in Hungary and X-ray imaging (X-ray) is used as a first line diagnostic method, in both COPD patient care and cancer screening, even though we know that X-ray is not suitable for detecting early-stage lung cancer. Routine diagnostics supplemented with a complete glycosylation analysis of the serum can be sufficiently informative about the condition of patients with a minimal increase in costs.

COPD, lung cancer and comorbid groups cannot be separated based on the degree of branching observed in the N-glycan profile of serum, so based on my results, the use of this feature as a biomarker for the effective classification of patient groups is not possible. Although I could not separate the comorbid patient group from either the control or the lung cancer groups with the statistical analysis of the data obtained by the N-glycosylation analysis of the serum, the above glycan structures may be disease-specific biomarkers, so their follow-up and further analysis is definitely recommended.

As I concluded from the SPM results that a significant part of the N-glycan structures showing statistically significant differences between the groups reflect the glycosylation status of the IgG antibodies circulating in the serum, it can be stated that the N-glycosylation analysis of the IgG molecule is useful and may be significant in the diagnosis of both lung cancer and various inflammatory pathologies.

My results show that N-glycomic analysis of serum samples from patients with inflammatory and neoplastic lung diseases provides useful and rapid preliminary molecular diagnostic information on global N-glycosylation changes and their potential glycoprotein sources.

5. The most important results of my research

1. Storage of the whole blood sample at room temperature for four and a half hours after blood sampling until centrifugation has no significant effect on the serum N-glycome, thus, serum can be considered a relatively stable source of glycobiomarkers.
2. Both maternal obesity and gestational diabetes affect the N-glycosylation of IgG molecules that transported across the placenta. This effect can be detected through the

sialylation of N-glycosylation of IgG molecules isolated from children a relatively stable source of glycobiomarkers.

3. IgG from healthy pregnant women passing through the placenta to the fetus has a lower degree of sialylation.
4. IgG is not a proper glycobiomarker for predicting GDM, the inflammation associated with it, or maternal complications that may arise from GDM.
5. On the other hand, the IgG N-glycan structures reflect the possible inflammation occurring in maternal obesity even during pregnancy.
6. Since the ratio of sialic acids terminating N-glycan structures of the IgG molecules was higher in children who was born to mothers diagnosed with obesity and GDM than the value observed in the children of the control group, it can be assumed that the placenta is able to adapt to the maternal inflammatory environment, protecting the fetus from the harmful inflammatory, oxidative, hyperinsulinemic environment present in maternal obesity and GDM.
7. Glycosylation-selective IgG transport may be part of this defense mechanism, and as a result the placenta may favor molecules with higher sialylation under inflammatory conditions. The reduced IgG sialylation observed in mothers with both disorders and in their children may indicate the limits of this protective mechanism, where the placenta can no longer to compensate for adverse conditions.
8. IgA glycosylation changes may serve as a disease-specific biomarker for GDM.
9. The SF/NF ratio of IgA derived structures was neither influenced by obesity, nor GDM and the combined presence of the two disorders in the case of the mothers. Therefore, IgA is not suitable for indicating the existence of inflammatory processes.
10. A glycoanalytical approach has been developed, applying the analysis of entire serum sample, it can provide information about the glycosylation modifications of the IgG and IgA antibodies present in serum, reducing the sample, material, and time required for protein analysis.
11. At the serum level, I identified several N-linked glycan structures, which successfully differentiated the COPD, lung cancer and comorbid patient groups from each other and from the control group.
12. IgG N-glycosylation can be a potential biomarker in the diagnosis of lung cancer and COPD, as a remarkable part of the structures showing significant differences between patient groups originate from IgG antibodies based on the developed model.

13. The N-glycomic analysis of human serum samples can provide quick, preliminary molecular diagnostics information about the global changes in serum N-glycome and their potential glycoprotein sources in relation to inflammatory and malignant lung diseases.

6. Summary

I investigated the possible effect of storage at room temperature (24-25°C) on the N-glycan composition depending on the time delay from blood collection to centrifugation and freezing using a CE-LIF device to determine the stability of serum N-glycome and to investigate the applicability of disease-induced changes as potential biomarker. As a result of my experiments, I found that the storage of whole blood at room temperature had no significant effect on the serum N-glycan profile. As a result of my work, it can be stated that the serum N-glycome can be considered a relatively stable source of biomarkers under the conditions I applied and could be routinely used in clinical laboratories.

Both obesity and GDM are occurring more and more frequently. Both obesity and GDM can have adverse effects on both the mother and the unborn child, given the greater number of cases there is a growing need to identify and uniformly apply biomarkers that may predict the development of GDM and its effects on the mother and her child. N-glycosylation of maternal and the child-derived serum IgG and maternal serum IgA has been shown to be a promising marker in my studies for the development and / or effects of obesity and / or GDM.
Based on my results, I concluded that the changes in the N-glycan structures of IgG excellently reflect the maternal obesity, but cannot indicate the development of GDM. In contrast, I have demonstrated in my work that IgA may be a potential source of glycobiomarkers in predicting the development of GDM.
Comparing the glycosylation of child and maternal IgG, I speculated that both maternal obesity and GDM can affect placental transport processes. In this context, I consider it appropriate to perform subclass and binding site-specific IgG glycosylation analysis with additional subjects, supplemented by analysis of inflammatory markers of the placenta. In this way, we can get closer to a better understanding of the transport processes of antibodies through the placenta and thus the factors influencing the development of the fetal immune system.

In the last part of my dissertation, a new approach (SPM model) was presented, which can be used to efficiently study the N-linked glycosylation modifications of the above mentioned two Ig classes in parallel with three other proteins present in high concentrations in human serum (haptoglobin, transferrin and alpha-1-antitrypsin).

Using the developed model, I examined the groups of patients with different inflammatory and neoplastic lung diseases in order to identify biomarkers, that can be used to differentiate the patient groups diagnosed with COPD, lung cancer, and both diseases. During my work, I determined the degree of antennary branching belonging to the desialylated N-glycan profiles of the serum samples of these groups, a property that allowed the differentiation of lung cancer from the control group.

In addition to the statistical analysis, I introduced a new evaluation procedure to examine the change in the relative area percentage of the identified structures. Using this method, I have been able to identify a number of glycan structures that may assist in diagnosing the disease in the future. Outstanding among these structures is the core-fucosylated biantennary, agalactosylated structure (FA2), which can effectively differentiate COPD from lung cancer and both lung diseases from the healthy control group.

Considering the results of the statistical analysis of each N-glycan structure, the SPM model suggested that N-glycosylation analysis of the IgG molecule may be of relevance for the diagnosis of lung cancer and various inflammatory diseases (COPD).

N-glycomic analysis of serum samples from patients with the lung diseases studied in this work may provide important additional and rapid preliminary molecular diagnostic information about global N-glycosylation changes and their potential glycoprotein sources, thereby facilitating future diagnosis and therapeutic decision-making.

7. References

1. Boerma, T.A.-Z., C; Kinfu, Y, *World health statistics 2008.* 2008: WHO Pres.

2. Deng, T., et al., *Obesity, Inflammation, and Cancer.* Annu Rev Pathol, 2016. **11**: p. 421-49.

3. Wilson, R.M. and I. Messaoudi, *The impact of maternal obesity during pregnancy on offspring immunity.* Mol Cell Endocrinol, 2015. **418 Pt 2**: p. 134-42.

4. Ygberg, S. and A. Nilsson, *The developing immune system - from foetus to toddler.* Acta Paediatr, 2012. **101**(2): p. 120-7.

5. de Jong, S.E., et al., *IgG1 Fc N-glycan galactosylation as a biomarker for immune activation.* Scientific Reports, 2016. **6**(1): p. 28207.

6. Barnes, P.J. and I.M. Adcock, *Chronic obstructive pulmonary disease and lung cancer: a lethal association.* Am J Respir Crit Care Med, 2011. **184**(8): p. 866-7.

7. Szigeti, M. and A. Guttman, *Sample Preparation Scale-Up for Deep N-glycomic Analysis of Human Serum by Capillary Electrophoresis and CE-ESI-MS.* Mol Cell Proteomics, 2019. **18**(12): p. 2524-2531.

8. Reider, B., M. Szigeti, and A. Guttman, *Evaporative fluorophore labeling of carbohydrates via reductive amination.* Talanta, 2018. **185**: p. 365-369.

9. Goonatilleke, E., et al., *Immunoglobulin A N-glycosylation Presents Important Body Fluid-specific Variations in Lactating Mothers.* Mol Cell Proteomics, 2019. **18**(11): p. 2165-2177.

10. Mesko, B., et al., *Peripheral blood gene expression and IgG glycosylation profiles as markers of tocilizumab treatment in rheumatoid arthritis.* J Rheumatol, 2012. **39**(5): p. 916-28.

11. Meszaros, B., et al., *Comparative analysis of the human serum N-glycome in lung cancer, COPD and their comorbidity using capillary electrophoresis.* J Chromatogr B Analyt Technol Biomed Life Sci, 2020. **1137**: p. 121913.

12. Secchiero, S., L. Sciacovelli, and M. Plebani, *Harmonization of units and reference intervals of plasma proteins: state of the art from an External Quality Assessment Scheme.* Clin Chem Lab Med, 2018. **57**(1): p. 95-105.

13. Varadi, C., C. Lew, and A. Guttman, *Rapid magnetic bead based sample preparation for automated and high throughput N-glycan analysis of therapeutic antibodies.* Anal Chem, 2014. **86**(12): p. 5682-7.

14. Jarvas, G., M. Szigeti, and A. Guttman, *Structural identification of N-linked carbohydrates using the GUcal application: A tutorial.* J Proteomics, 2018. **171**: p. 107-115.

15. Friedman, M., *The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance.* Journal of the American Statistical Association, 1937. **32**(200): p. 675-701.

16. Sinkovits, G.P.Z. *Paraméteres és nem paraméteres próbák alkalmazása több csoport összehasonlítására folytonos változók esetén*. 2021; Available from:

https://semmelweis.hu/kutlab/files/2021/03/Csoportok_%C3%B6sszehasonl%C3%ADt%C3%A1sa_2021_SGy.pdf.

17. Dunn, O.J., *Multiple Comparisons Using Rank Sums.* Technometrics, 1964. **6**(3): p. 241-252.

18. Anderson, N.L. and N.G. Anderson, *The human plasma proteome: history, character, and diagnostic prospects.* Mol Cell Proteomics, 2002. **1**(11): p. 845-67.

19. Saldova, R., et al., *Ovarian cancer is associated with changes in glycosylation in both acute-phase proteins and IgG.* Glycobiology, 2007. **17**(12): p. 1344-56.

20. Kovacs, Z., et al., *Capillary electrophoresis analysis of N-glycosylation changes of serum paraproteins in multiple myeloma.* Electrophoresis, 2017. **38**(17): p. 2115-2123.

21. Mariño, K., et al., *Changes in serum N-glycosylation profiles: Functional significance and potential for diagnostics.* Carbohydrate Chemistry, 2011. **37**: p. 57-93.

22. Arnold, J.N., et al., *Novel glycan biomarkers for the detection of lung cancer.* J Proteome Res, 2011. **10**(4): p. 1755-64.

23. Dedova, T., et al., *The effect of blood sampling and preanalytical processing on human N-glycome.* PLoS One, 2018. **13**(7): p. e0200507.

24. Denison, F.C., et al., *Obesity, pregnancy, inflammation, and vascular function.* Reproduction, 2010. **140**(3): p. 373-85.

25. Ruhaak, L.R., et al., *Total plasma N-glycome changes during pregnancy.* J Proteome Res, 2014. **13**(3): p. 1657-68.

26. Štambuk, T., et al., *Associations between plasma protein, IgG and IgA N-glycosylation and metabolic health markers in pregnancy and gestational diabetes*. 2021.

27. Raju, T.S. and S.E. Lang, *Diversity in structure and functions of antibody sialylation in the Fc.* (1879-0429 (Electronic)).

28. Bondt, A., et al., *Association between Galactosylation of Immunoglobulin G and Improvement of Rheumatoid Arthritis during Pregnancy Is Independent of Sialylation.* Journal of Proteome Research, 2013. **12**(10): p. 4522-4531.

29. Williams, P.J., et al., *Short communication: selective placental transport of maternal IgG to the fetus.* (0143-4004 (Print)).

30. Steffen, U., et al., *IgA subclasses have different effector functions associated with distinct glycosylation profiles.* Nature Communications, 2020. **11**(1): p. 120.

8. List of publications

UNIVERSITY of DEBRECEN

UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

Registry number: DEENK/369/2022.PL
Subject: PhD Publication List

Candidate: Anna Farkas
Doctoral School: Doctoral School of Molecular Medicine

List of publications related to the dissertation

1. Török, R., **Farkas, A.**, Guttman, A., Járvás, G.: Evaluation of Possible Processing Time Effects on the Global N-Glycosylation Profile of Human Blood Samples.
   Curr. Mol. Med. 20 (10), 840-846, 2020.
   DOI: http://dx.doi.org/10.2174/1566524020666201230094722
   IF: 2.222
2. **Farkas, A.**, Mészáros, B., Szarka, M., Szigeti, M., Kappelmayer, J., Szabó, M., Csánky, E., Guttman, A.: Modeling of the Desialylated Human Serum N-glycome for Molecular Diagnostic Applications in Inflammatory and Malignant Lung Diseases.
   Curr. Mol. Med. 20 (10), 765-772, 2020.
   DOI: http://dx.doi.org/10.2174/1566524020666200422085316
   IF: 2.222

UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

List of other publications

3. Mészáros, B., Járvás, G., **Farkas, A.**, Szigeti, M., Kovács, Z., Kun, R., Szabó, M., Csánky, E., Guttman, A.: Comparative analysis of the human serum N-glycome in lung cancer, COPD and their comorbidity using capillary electrophoresis.
J. Chromatogr. B. 1137, 1-7, 2020.
DOI: http://dx.doi.org/10.1016/j.jchromb.2019.121913
IF: 3.205

Total IF of journals (all publications): 7,649
Total IF of journals (publications related to the dissertation): 4,444

The Candidate's publication data submitted to the iDEa Tudóstér have been validated by DEENK on the basis of the Journal Citation Report (Impact Factor) database.

14 July, 2022

DEBRECENI EGYETEM
Egyetemi és Nemzeti Könyvtár