

Hybrid Mass Spectrometric Approaches Expose the Diverse Proteoform Profiles of Serum Glycoproteins

Yu-Hsien Lin

林育賢

ISBN: 978-94-6416-253-0

ELEKTRONIC ISBN: 978-94-6416-254-7

The research presented in this thesis was performed in the Biomolecular Mass Spectrometry & Proteomics Group at the Bijvoet Center for Biomolecular Research, Utrecht University, The Netherlands.

Copyright © Yu-Hsien Lin

Printed by Ridderprint | www.ridderprint.nl

Hybrid Mass Spectrometric Approaches Expose the Diverse Proteoform Profiles of Serum Glycoproteins

**Hybride massaspectrometrische methoden onthullen de diverse proteovorm
profielen van geglycosyleerde serumeiwitten**

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

maandag 30 november 2020 des ochtends te 11.00 uur

door

Yu-Hsien Lin

geboren op 5 december 1990
te Taichung, Taiwan

Promotor:

Prof. dr. A.J.R. Heck

Table of contents

Chapter 1	7
Introduction to protein glycosylation	
Chapter 2	45
Similar albeit not the same; in-depth analysis of proteoforms of human serum, bovine serum and recombinant human fetuin	
Chapter 3	71
Glycoproteogenomics: a frequent gene polymorphism affects the glycosylation pattern of the human serum fetuin/ α -2-HS-glycoprotein	
Chapter 4	113
Glycoproteoform profiles of individual patients' plasma α -1-antichymotrypsin are unique and extensively remodeled in response to a septic episode	
Chapter 5	149
A quantitative longitudinal inventory of the N-glycoproteome of human milk from a single donor reveals the highly variable repertoire and dynamic site-specific changes	
Chapter 6	183
Neutrophil azurophilic granule proteins are distinctively decorated by highly atypical pauci- and phosphomannose glycans	
Chapter 7	215
Summary and Outlook	
Appendix	229
Curriculum Vitae	
List of publications	
Acknowledgments	

Chapter 1

Introduction to protein glycosylation

1. Introduction to protein glycosylation

Glycosylation is one of the most important post-translational modifications (PTMs), which occurs on many proteins^{1,2}. Glycosylation involves the covalent attachment of varying structures of carbohydrates (glycans) to specific amino acid residues of polypeptide chain. There are two common types of protein glycosylation (**Figure 1**): N-linked glycosylation, which refers to the attachment of glycans to proteins via the side-chain nitrogen of asparagine (Asn), and O-linked glycosylation, which refers to the attachment of glycans to proteins via the side-chain oxygen of serine or threonine. All N-linked glycans share a common core structure, containing two N-acetylglucosamines (GlcNAc) and three mannoses (Man), which is covalently bound to a consensus sequence N-X-S/T motifs (sequones), where X can be any amino acid except proline. Enzymatic reactions such as galactosylation, sialylation, and fucosylation can add a variable number of other monosaccharides to the core structure. The sequence of these reactions results in three main types of N-glycans (**Figure 1**): high mannose, hybrid, and complex.

Unlike N-linked glycosylation, which is driven by a canonical motif and a precursor glycan is transferred *en bloc* to Asn, O-linked glycosylation is synthesized in a stepwise fashion by adding glycan units incrementally on the serine and threonine residues. A consensus sequence for the initiating glycan addition has not been found, although predictive algorithms do exist and can help with discovering new O-glycosylation sites³. The most common O-glycosylation is called mucin-type O-glycosylation. It involves adding monosaccharide, typically N-acetylgalactosamine (GalNAc). The GalNAc can be further elongated with various glycan units by a protein family of 20 distinct galactosyltransferases (GALNTs). These enzymatic reactions result in the production of eight core structures (core-1 to core-8), of which core-1 and core-2 (**Figure 1**) are the main types of O-glycans that occur on human plasma proteins. In addition to the mucin-type of O-glycosylation, there is another well-known type of O-glycosylation, O-GlcNAcylation, which has a fundamentally different biological purpose. In this case, GlcNAc group can be added or removed dynamically from serine or threonine and play a role in signaling pathways⁴. Other types of O-glycosylation, such as O-linked Man and O-linked fucose, often occur on specific protein domains, such as epidermal growth factor (EGF) repeats and thrombospondin type 1 repeats (TSR)⁵. For the sake of completeness, there are also rare types of glycosylation that should be mentioned. One is C-mannosylation, which is a modification of tryptophan (W)⁶, and the other is S-glycosylation, where glycans are added to the sulfur atoms of cysteine residues⁷.

N- and O-linked glycans are crucial for protein functioning and ensuring the proper folding, conformation, stability, and play roles as ligands in inter-/intra-cellular recognition and host-pathogen interaction⁸. Further, the glycans can modulate various biological processes, such as cell growth, receptor activation, signal transduction, endocytosis, and viral replication⁹. Such a broad repertoire of functional abilities is mediated through the enormous structural complexity of glycans. Understanding this complexity requires a deep knowledge of biological processes involved in glycosylation pathways and the development of new and more sophisticated methods to characterize glycoprotein structure and function. The advancements and developments in glycobiology go hand-in-hand with improving human health and medicine. Protein glycosylation has been shown to be involved in pathologies of many diseases, including infection, genetic disorders, developmental defects, and cancer⁵. Understanding glycosylation function also leads to essential improvements in applying protein drugs. One example is the monoclonal antibody cetuximab, which has been used in cancer immunotherapy for more than 20 years. At first, approximately 3% of patients suffered from severe immune hypersensitivity reactions to the treatment by cetuximab¹⁰. A few years later, scientists found out that these undesired immune reactions were associated with IgE antibodies against galactose- α -1,3-galactose (alpha-gal allergy). This glycan is present on the Fab portion of the heavy chain of cetuximab, which is produced in the mouse myeloma cell line¹¹. Therefore, glycosylation is strictly monitored during the protein

drug production process, and all recombinant chimeric, humanized, or human therapeutic antibodies are produced in cells that can synthesize glycosylation patterns with favorable bioactivity and improved acceptance in patients^{12–15}.

Development and production of new and better glycosylated monoclonal therapeutics is not the only way improved knowledge about glycosylation can be utilized. Deep understanding of the biology of glycosylation can also be used as a start to manipulate biochemical pathways in the cells, which can have applications in basic research and therapy in general. This was nicely demonstrated in a study from Bannik et al., where the authors repurposed the protein degradation pathway for directing secreted and membrane proteins for degradation. In this strategy, the known mannose-6-phosphate pathway was utilized to degrade targeted proteins, such as the epidermal growth factor receptor (EGFR), Apolipoprotein-E4, CD71, or programmed death-ligand (PD-L1), which allowed to study various biological processes, including receptor trafficking and protein degradation¹⁶.

A large part of medicinal and clinical research is focused on analyzing and understanding processes in blood and serum. One of the major challenges in glycobiology has been for a long time the analysis of human serum glycoproteins. More than 50% of serum proteins have been found to be glycosylated¹⁷. Human serum is a source of many potential biomarkers for various diseases. Recent advances in mass spectrometry methods and automated data processing tools have allowed comprehensive and large-scale analysis of the human serum glycoproteome^{18,19}. Described changes in glycosylation of serum glycoproteins in healthy people and patients can reflect their physiological and pathological states and lead to the discovery of new biomarkers and medicines²⁰.

This thesis focuses mainly on protein N-glycosylation and describes its biological relevance and analysis. The first introductory part describes the biosynthetic pathway of N-glycans, including the synthesis of the main glycosylation types occurring in humans. I touch upon N-glycosylation diversity and introduce some of the important features of glycan structures. Next, I explain the importance of the human plasma proteome and plasma glycoproteins in the context of its use in clinical research as biomarkers. In the end, I discuss a variety of disease-related glycosylation and pathological mechanisms related to aberrant or altered glycosylation. In the second part of the Introduction, I mention some strategies for N-glycosylation analysis, including glycomics and glycoproteomics analytical approaches. Mainly, I focus on enrichment and separation methods for the analyses of N-glycans, intact N-glycopeptides, and intact glycoproteins. I highlight the role of mass spectrometry methods and describe in detail the benefit of combining different methods for the characterizing of glycoproteins. In Chapter 2 I compared fetuin proteins derived from three biological sources—human, cow, and recombinant cellular fetuin—to show how the expression of a single gene in different species can result in extremely different proteins with different primary structures, glycosylation, and phosphorylation profiles. In Chapter 3 I compared proteoform profiles of fetuins purified from the blood serum of healthy people and septic patients. Here, I discuss altered glycosylation in septic patients compared to healthy ones and show the importance of considering patient genotype for glycosylated biomarkers. In Chapter 4 I present a longitudinal study performed on a cohort of septic patients, showing the extensive remodeling of glycosylation of serum acute phase glycoproteins across different time points. In Chapter 5 I present an efficient, automated HILIC-based platform, which enables the efficient and robust enrichment of human milk N-glycopeptides. Finally, in Chapter 6 I applied a HILIC-based stage-tip approach to analyze in-depth the neutrophil glycoproteome. This chapter mainly focuses on neutrophil proteins modified by paucimannosidic glycans.

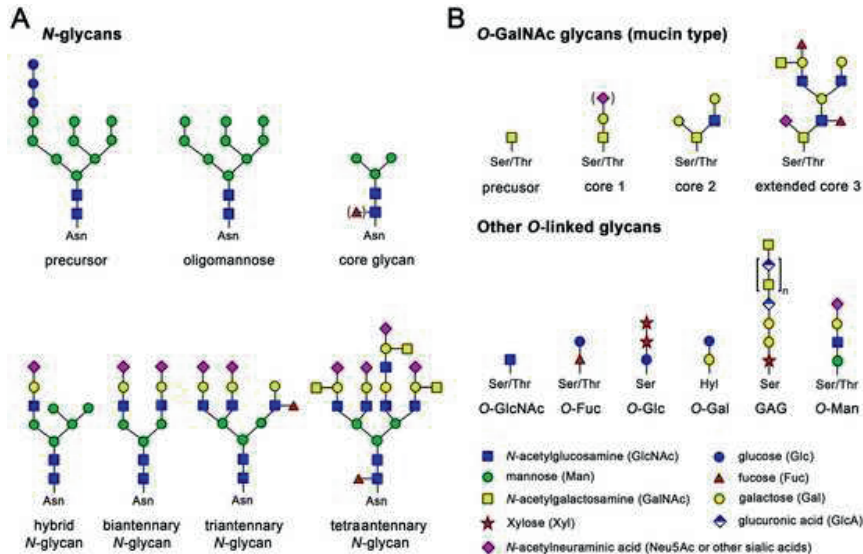


Figure 1. Major types of glycosylation in humans. Depiction of different types of glycans, including N-linked (high mannose, hybrid, complex type), O-linked (core-1, core-2), and O-GlcNAc. Figure adapted with permission from Goettig²¹.

1.1. Biosynthetic pathway of N-glycans

The biosynthetic pathway of N-glycans (**Figure 2**) can be divided into three stages: an assembly of lipid-linked oligosaccharide, transfer of oligosaccharide to protein, and processing of oligosaccharide. The first two stages occur in the endoplasmic reticulum (ER) and the last stage in the Golgi apparatus. N-linked protein glycosylation is initiated by assembling a $\text{Man}_5\text{GlcNAc}_2$ structure to the dolichol phosphate (Dol-P) through a pyrophosphate linkage at the cytoplasmic side of the ER membrane. A series of glycosyltransferases catalyzes the addition of two N-Acetylglucosamine (GlcNAc) and five Mannose (Man) residues derived from two nucleotide-activated sugars (UDP-GlcNAc and GDP-Man). This 7-sugar glycan is then translocated across the ER membrane and exposed to the luminal side of the ER. The growing glycan chain is further extended with four Man and three glucose (Glc) residues derived from Dol-P-linked Man and Dol-P-linked Glc, forming a $\text{Glc}_3\text{Man}_4\text{GlcNAc}_2$ oligosaccharide structure in the lumen of the ER. The critical step in the ER is the transfer of the glycan from the Dol-P-linked carrier to the asparagine (Asn) residues within the respective N-X-S/T sequons (X should not be P) in the growing protein regions. The primary enzyme that catalyzes this reaction in the luminal side of the ER is the oligosaccharyltransferase (OST). Although OST shows high ability to add glycans on the glycosylation sequons, roughly one-third of the potential sites are left unoccupied. Several factors affect the OST activity. First, OST is a substrate-specific enzyme, showing a higher preference for the NXT sequon over the NXS sequon, which was found in the major model organisms such as *Caenorhabditis elegans* and *Danio rerio*²². Second, the glycan transfer depends on the local conformation of the newly synthesized polypeptide. More specifically, it has been shown that many turns or loop conformations may contain an exposed N-glycosylation site for the glycan transfer²³, and 78% of N-glycosylation sites are situated in the loop/turn conformation in the human proteome²⁴. In summary, the attachment of N-glycan on proteins

requires strict concordance with the consensus motifs, OST specificity, and topological location of the substrates.

Before the newly synthesized N-glycoprotein reaches the final stage of the biosynthetic pathway in the Golgi, many processing steps are taken in the ER, which play critical roles in regulating glycoprotein folding and quality control system. The processing in the ER is mainly defined by the action of glucosidases, which trim Glc residues. Eventually, these processes lead to the formation of high mannose type glycans, usually containing eight or nine Man residues²⁵. Glycoproteins with these N-glycans may directly passage through the Golgi without changing compositions and be secreted to the cell surface. Nevertheless, high mannose structures are often processed in the cis-Golgi by a series of α -mannosidases, which gives a key intermediate Man:GlcNAc₂ in the pathway to hybrid and complex N-glycans. Synthesis of these glycans is initiated in medial-Golgi by GlcNAc-transferase 1, which transfers one GlcNAc residue to the arm of the glycan structure with fewer Man residues²⁵. Next, mannosidases remove two Man residues to form GlcNAcMan:GlcNAc₂ and a second GlcNAc is added. This series of reactions results in the precursor for all complex N-glycans²⁵. The complex N-glycans commonly have two branches, but it is possible to add up to 6 GlcNAc residues and yield tri- and tetra-antennary glycans. Each antenna may be further elongated by adding other sugars such as galactose (Gal), fucose (Fuc), and sialic acid (SA). Hybrid glycans are formed in case mannosidase does not remove Man residues from GlcNAcMan:GlcNAc₂. Hybrid glycans can keep the five Man residues and extend the arm that received GlcNAc by Gal, Fuc, and SA²⁵. Recent studies have assembled evidence that a fourth type of protein N-glycosylation, referred to as paucimannosylation, is present in humans²⁶⁻²⁸. Paucimannose type of glycosylation has compositions of Man₁₋₃:Fuc₀₋₁:GlcNAc₃, but the biosynthetic pathway of these glycans still needs to be clarified.

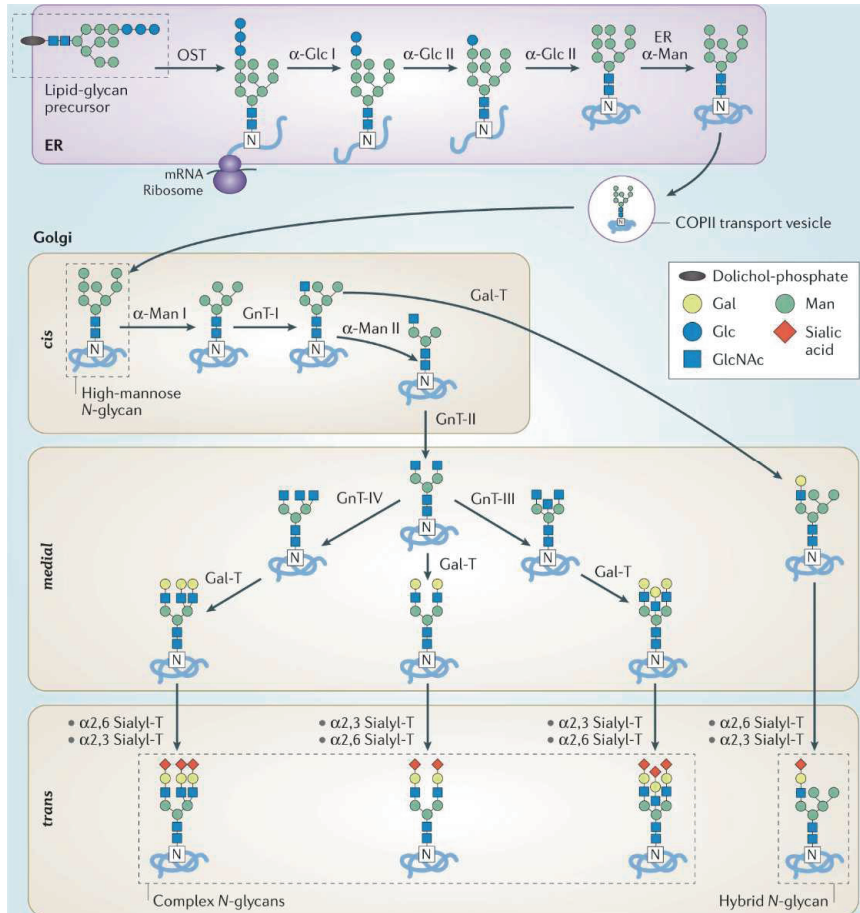


Figure 2. N-glycan biosynthesis in the secretory pathway. N-glycan synthesis begins with a co-translational en bloc transfer of a lipid-glycan precursor (Dolichol-phosphate-Glc₃Man₆GlcNAc₂) to Asp by the enzyme oligosaccharyltransferase (OST). The processed structure moves from the ER to the cis-Golgi apparatus, where the carbohydrate structures are trimmed further. Maturation occurs in the trans-Golgi apparatus. N-glycans can be capped with monosaccharide residues by various glycosyltransferases and are finally transported in vesicles to the cell surface. Figure adapted with permission from¹.

1.2 N-glycan Diversity and Heterogeneity

In 2005, Kawano et al. statistically analyzed the KEGG database (Kyoto Encyclopedia of Genes and Genomes)^{29,30} to predict glycan structures from gene expression data. The authors listed 4107 unique glycan structures, which mostly consisted of 9 common monosaccharides (N-acetylglucosamine, N-acetylgalactosamine, glucose, galactose, mannose, fucose, xylose, glucuronic acid, and sialic acid). The data showed that only 302 out of 558 theoretically possible disaccharides (9 monosaccharides, 2 anomers, 31 substitution possibilities) were present in the database. Moreover, there is an enormous number of theoretically possible combinations for

glycosyltransferase-reaction schemes, but only around 2000 of them existed in the database. These observations indicate that although the structural diversity of glycans is genuinely enormous, the glycosylation machinery in the cellular environment does not produce just all statistically possible glycan combinations but is limited by the ability of enzymes involved in the biosynthesis to build up these structures. The explanation of glycan diversity lies in the deeper understanding of glycosylation biosynthetic pathway and biochemical process related to the glycosylation in general.

Glycosylation biosynthesis is a non-template driven process in the ER and Golgi. It involves a variety of enzymatic reactions, which are not always 100% efficient, leading to incomplete addition of glycans to proteins. Moreover, the competition among multiple glycosyltransferases for the same glycan substrates may also cause incomplete processing of the glycans^{31,32}. These processes lead to the high structural diversity of glycans. The structural diversity of N-glycans can be classified into two classes: 1) the macroheterogeneity, which results from the presence or absence of glycan at a specific site in a protein, and 2) the microheterogeneity, which results from the presence of various glycan structures at a specific glycosylation site³³. As mentioned in Chapter 1.1, three common N-glycosylation classes (high mannose, complex, and hybrid) are formed by various glycosidases and glycosyltransferases (GnTs) in Golgi^{34–36}. In the following paragraph, I will focus on three important GnTs, GnT-III, GnT-V, and Fut8, which are involved in the synthesis of three biologically active glycan sub-types: bisecting glycans, N-acetylglucosaminyl (GlcNAc) glycans, and core-fucosylated glycans. I selected these glycan subtypes due to their association with various diseases, which I will further describe in Chapter 1.4.

Complex and hybrid N-glycans may be modified by the action of GnT-III, which generates “bisecting” N-glycans¹⁸¹ (**Figure 3a, b**). Particularly, GnT-III catalyzes the addition of a GlcNAc residue to the β -Man residue of the N-glycan core structure. This bisecting GlcNAc cannot be further modified, unlike GlcNAc forming branches^{37,38} (**Figure 3c**). Moreover, the presence of a bisecting GlcNAc prevents further glycan extension by other GnTs responsible for branching³⁹. With this ability, the upregulation of GnT-III may potentially influence the production of other types of N-glycan structures. Therefore, the final appearance of glycan structures on the synthesized glycoproteoforms significantly depends on the interplay between GnT-III and other GnTs.

The second functional modification of complex and hybrid glycans is the addition of LacNAc. In this case, GnT-V catalyzes the addition of a GlcNAc residue to the α 1,6-Man and forms β 1,6-branched glycans¹⁸¹ (**Figure 3c**). This GlcNAc is often extended by the addition of β -linked Gal, forming a LacNAc (Gal β 1-4GlcNAc) building block. The LacNAc antennae can be further elongated by the sequential addition of LacNAc blocks, forming a poly-LacNAc structure⁴⁰. Poly-LacNAc N-glycans serve as high-affinity ligands to galectins. Galectin-glycan interactions are known to be involved in the signaling pathway of growth factor receptors at the cell surface⁴⁰. Dysregulation of these processes may contribute to aberrant immune reactions and autoimmune diseases⁴¹.

The third modification in line is the main core fucosylation mediated by fucosyltransferase, Fut8. Fut8 catalyzes the addition of a Fuc residue to the initial GlcNAc residue of N-glycan core¹⁸¹ (**Figure 3a, b**). Core fucose structure has clinical relevance as a biomarker for the detection of cancer⁴². Other fucosyltransferases (Fut3, Fut4, Fut5, Fut6, Fut7, and Fut9) synthesize α 1,3-, and in the case of Fut3 α 1,4-fucosylated glycans, such as the Lewis^x (Le^x) and sialyl Lewis^x antigen⁴³. Also, these modifications are associated with cancer biology⁴³, and I will discuss it more in detail in Chapter 1.4.

There are many reasons behind the structural diversity of glycans in nature, such as the remarkably complex biosynthetic pathway or structural variability of monosaccharides. Nevertheless, deeper explanation and understanding why this complexity exists and what is the role of glycans in cellular processes and molecular interactions still remain a subject of scientific discussions and theories. One of the well-accepted hypothesis, the “Red Queen” effect, proposes that species must continuously adapt, develop, and proliferate in order to survive while exhibited to constantly evolving opposing species⁴⁴ (Figure 3d). If we consider that all cells are covered with glycans, which are important not only for the functioning of hosts but also for interacting with pathogens, the application of this evolutionary theory can very well explain the structural diversity of glycans. The surface glycans of host cells are often targets of bacterial and viral infections. Hosts survive by continually changing the structures of surface glycans without affecting their cellular functions, preventing pathogens from binding to the glycans. However, glycan-binding pathogens have short generation times and high mutation rates, allowing them to evolve much more rapidly than their hosts with longer life cycles. Eventually, this generates an evolution cycle between hosts and pathogens. Evolutionary processes drive the potential for variability provided by glycosylation machinery and generate the enormous complexity in glycan structures. The Golgi is well-suited for hosts to initiate this cycle because of its non-template-driven and assembly line-like mechanisms for creating widely diverse glycans. In conclusion, the hallmark of protein N-glycosylation is in the production of highly diverse glycan structures.

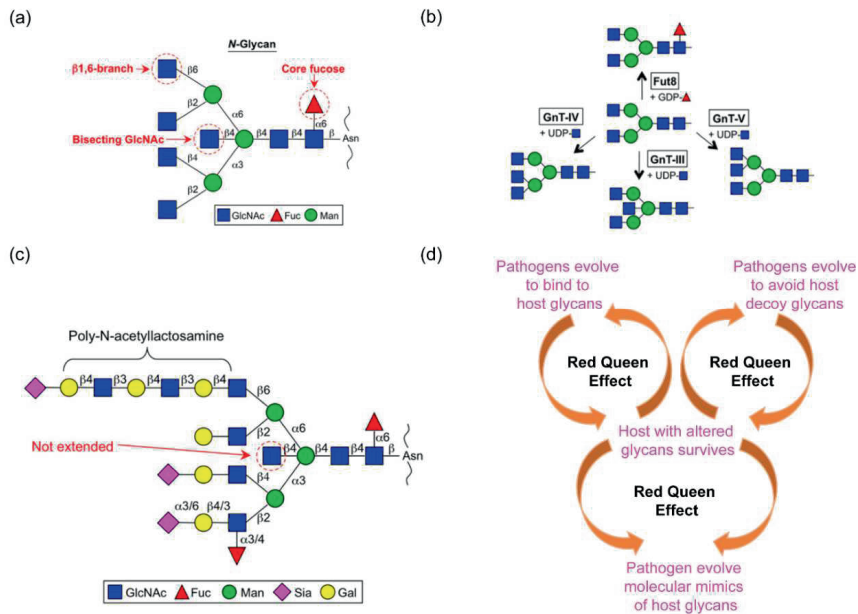


Figure 3. (a) Structure of N-glycan with core-fucosylation and different GlcNAc branches, including bisecting GlcNAc branch. (b) Specific enzymes responsible for the biosynthesis of N-glycan branches and core-fucosylation. (c) Bisecting GlcNAc (indicated by a red dashed circle) is not further added to other glycan residues, whereas other GlcNAc branches are usually elongated by Gal and Sia. The β 1,6 GlcNAc branch can be elongated with a number of N-acetylglucosamine (LAcNAc) residues (poly-LAcNAc). (d) Red queen effect: each arrowed circle shows the interaction between host and pathogen. Hosts are constantly trying to evolve in order to prevent pathogens from binding, while pathogens are also constantly adjusting to changes made by hosts. Figure adapted with permission from^{44,45}.

1.3. Importance of the Serum Proteome and in particular Serum Glycoproteins

Human plasma is composed of a diverse set of proteins that carry out various physiological functions and play a central role in facilitating diverse biological processes. In 2017, the Human Plasma Proteome Project (HPPP) had cataloged 3500 serum proteins that have met the Human Proteome Project Data Interpretation Guidelines⁴⁶. All these proteins are very variable in their function, structure, abundance, and many other parameters, and therefore, it is challenging to classify them based on a few simple criteria. One classification can be based on the distinct origin of various plasma proteins. There are roughly four different categories: 1) proteins secreted by the liver; 2) immunoglobulins; 3) proteins released from cells to the plasma due to cell death or damage; 4) proteins released from tumors and other disease-affected tissues⁴⁷. Another common classification can be based on the protein's functionality. The first class contains proteins with a functional role in blood (human serum albumin, apolipoproteins, acute-phase proteins). The second class are tissue leakage proteins. Their function is either not known or is not well-defined in the circulation. The third class are low abundant signaling molecules (such as protein hormones, cytokines).

Traditionally, plasma is an often-used source for diagnosis in the clinic. This is mainly due to the possibility of obtaining rich information reflecting an individual's health state, which is minimally invasive and cost-effective. In the daily clinical practice, the quantitative analysis of concentrations of many plasma components such as C-reactive protein (CRP)⁴⁶ and serum amyloid A⁴⁸ is routinely performed using single-protein immunoassays, monitoring, for instance, acute and chronic inflammation. In addition to these immunoassays, MS-based plasma proteomics has regained interest in the clinic due to recent technological advances in sensitivity and throughput^{49,50}. MS can provide the high-throughput and quantitative measurements of many plasma proteins in parallel and potentially discover protein biomarkers in larger patient cohorts. However, in plasma proteomics, many challenges remain to be addressed. One major problem is the presence of a few highly abundant proteins that dominate the total protein content; about 55% of the total plasma proteins consists of albumin, and together with other seven proteins (e.g., transferrin, immunoglobulins), they make up 85% of the total protein mass⁵¹. Although commercially available depletion columns can immunodeplete up to the 20 most abundant proteins in plasma, the depletion process can also induce a bias related to cross-reactions of used antibodies or proteins bound to carrier proteins like albumin^{52,53}. Another challenge is a lack of robust and high-throughput proteomics workflows that can reproducibly cover the plasma proteome with the acquired depth. Just to give an idea, the Plasma Proteome Database⁵⁴ (last update 2014) contains 10,546 proteins linked to 509 scientific articles. However, the detection of 2378 proteins of these is supported by two or more peer-reviewed publications, for all other 8000 very limited experimental evidence is available.

To address these challenges, further development of the plasma proteomics pipeline is essential, improving throughput, depth of proteome coverage robustness, and applicability of the underlying workflow. Recently, Geyer et al. developed an automated MS-based workflow that increased throughput, also shortening the time of analysis to less than 3-hr per sample⁵⁵. The authors used this method to perform a longitudinal study of 43 patients over 12 months and detected, on average, 437 proteins per person⁴⁹. Pernemalm *et al.* subsequently presented a method for the in-depth analysis of the human plasma proteome by using high-resolution isoelectric focusing HiRIEF LC-MS/MS. They achieved high plasma proteome coverage and demonstrated high reproducibility and sensitivity. The method additionally showed a potential for liquid biopsy protein profiling due to its ability to detect low abundant tissue-annotated proteins and single amino acid variants⁵⁶. The ultimate goal of in-depth plasma analysis for application in human medicine still requires much effort, including incorporating broader technical performance characterization into the method development and understanding

standardization and application of analytical processes. It is well known that any variation in the collecting and processing of plasma samples from blood, prior to the analytical phase, may cause a systematic bias^{57,58}. Therefore, the scientific community involved in human plasma research developed detailed sample preparation guidelines, which should help to prevent or minimize overall sample-related biases (see www.plasmaproteomeprofiling.org).

Apart from the standardization of the sample preparation procedures, one of the significant steps forward in the analysis of biomarkers in human plasma would include the characterization of protein post-translational modifications. As mentioned above, plasma proteins are variable in many different aspects, and PTMs significantly contribute to this complexity. Interestingly, most cancer protein biomarkers approved by the U.S. Food and Drug Administration (FDA) are heavily glycosylated⁵⁹. Aberrant glycosylation is a known hallmark of cancer and represents a promising diagnostic feature for monitoring disease prognosis and pathophysiology. Unfortunately, the prognostic capability of most protein biomarkers remains controversial due to the lack of specificity and sensitivity. One of the most valuable groups of serum proteins for monitoring pathological processes in the human body are acute-phase proteins (APPs). The innate acute phase response is a complex systemic early defense system triggered by different stimuli, including tissue injury, trauma, infection, and inflammation. APP proteins undergo quantitative and structural changes caused by inflammation related processes⁶⁰. The most common APPs include α -1-acid glycoprotein (AGP), haptoglobin (HPT), α -1-antitrypsin (AACT), α -fetoprotein (AFP), and fetuin (FET). The liver produces all these proteins, and all of them are highly glycosylated. Changes in glycosylation of APPs such as increased fucosylation and branching have been associated with the development and progression of various disease states such as cancer and acute inflammation⁶¹. For example, a decrease of tri-antennary structures in serum AGP was observed in sera of pancreatic cancer patients compared to controls, whereas a decrease of bi-antennary structures of AGP was shown in chronic pancreatitis patients compared to controls⁶². Another example by Wang et al. demonstrated a significant correlation between increased fucosylation levels of serum AFP (AFP-L3) and hepatocellular carcinoma (HCC), allowing to distinguish these patients from the ones suffering from chronic liver disease⁶³.

While the analysis of changes in the glycosylation of serum glycoproteins may represent a promising quantitative strategy for monitoring the disease status of patients, data interpretation is still very challenging due to the multifaceted features of protein glycosylation and limitations in the research methodology to date. One of the issues to be tackled is in the accurate quantitative analysis of the glycan changes on individual glycoproteins. Due to the dynamic changes in the relative abundance of the proteins during inflammation, it is not always apparent whether the change in glycosylation originates from a specific glycan alternation or a change in relative protein abundance⁶⁴. Moreover, the measurement of protein abundance is hampered by frequent genetic polymorphism of proteins, which may also influence frequently used clinical quantitative assays such as enzyme-linked immunosorbent assay (ELISA)⁶⁵. Many serum glycoproteins such as HPT⁶⁶ and FET⁶⁷ are highly polymorphic, resulting in different protein variants with different backbone sequences caused by one or a few amino acid substitutions. Genetic polymorphism may affect not only the protein concentration levels in plasma⁶⁵ but also the PTMs on certain sites of these proteins²⁰. In the case of HPT, the genetic polymorphism has even consequences on the higher-order structure of the protein, when different HPT genotypes show different oligomeric states of HPTs⁶⁸. In this context, a real understanding of glycosylation requires the systems-wide study, considering genotype, dynamics of protein serum level, N-glycosylation occupancy, and N-glycan compositions at specific sites⁶⁹. Together, this poses still an important challenge that I have tried to address in work described in this thesis.

1.4 Disease-associated changes in N-linked glycosylation

N-linked glycosylation exerts a fundamental role in numerous physiological and pathological processes, such as inflammation, infection, and cancer. Aberrant expression of N-linked glycoproteins has been linked to various disease phenotypes. Not surprisingly, several studies have been reported showing that changes in protein glycosylation can be utilized as biomarkers for disease diagnosis^{70,71}. These studies are built upon the hypothesis that the glycosylation of a given protein secreted from diseased tissues or malignant cells differ from the healthy control⁷². Over the past decade, an impressive compilation of disease-associated changes in N-glycans have been reported, related to diseases like breast cancer⁷³, prostate cancer⁷⁴, hepatocellular carcinoma (HCC)⁷⁵ and sepsis⁷⁶. Here, I will summarize the current knowledge of alternations in N-glycans related to specific disease states, categorized into changes in I. sialylation, II. fucosylation, and III. glycolytic branching.

I. In protein N-glycosylation, a Sialic acid often represents the terminal monosaccharide linked to the other monosaccharide residues, such as galactose, possibly by either one of three types of linkages: α 2,3-, α 2,6- or α 2,8-. As the Sialic acid is at the terminal location and harbors a negative charge, it often plays a role in regulating the interaction between these glycans and other ligands in the proximity of the glycoprotein. Over the years, to predict the state of a pathological process, changes in the level of sialylation have been measured either in specific glycoproteins or glycoproteome-wide the total amount of sialic acids present in tissue or body fluids. For example, several studies have revealed that the total level of sialic acid in human plasma increases rapidly upon inflammation^{77,78}, probably resulting from increased levels of sialylated acute-phase glycoproteins, secreted from hepatocytes. Related to this, Piagnerelli et al. demonstrated that sialylation in the plasma protein transferrin (TRF) is altered rapidly in sepsis⁷⁹. The proportion of N-glycan structures containing sialic acids (di-, tri- and tetraantennary) dramatically increased, even within 24 h, in sepsis patients when compared with healthy people. Also, the distinct linkage-isomers of sialyl moieties are involved in various biological processes. It has been reported that α 2,3-linked sialic acids are associated with malignancy in several types of cancer^{74,80,81}. The α 2,6-linked sialic acids are similarly linked to cancer, as it is presumed that they can serve as a regulator for blocking galectin binding⁸²⁻⁸⁴ and promote tumor cells growth⁸⁵. Although sialylation can be linked to disease states, it remains one of the more challenging aspects of analysis in glycoproteomics. In mass spectrometry, the loss from the protein or the peptide of the labile sialic acid groups and concomitantly a decrease of sialoglycopeptide ionization efficiency hamper such analyses. Also, the mass spectrometric distinction between linkage isomers of sialic acids remains a considerable challenge.

II. Fucosylation of glycoproteins is regulated by a variety of fucosyltransferases, which can result in the synthesis of several types of fucosylated glycan structures. Among these, sialyl Lewis-x, sialyl Lewis-a and α 1,6-linked fucosylation (core fucosylation) represent in human cells the most prominent fucosylated glycans. Growing evidence has indicated that increased fucosylation can be a valuable indicator of liver disease progression in HCC. For example, increased core fucosylation of the serum alpha-fetoprotein (AFP) has been observed in HCC patients and has been clinically used as a FDA-approved biomarker for distinguishing HCC from chronic liver disease⁷⁵. Furthermore, Gornik et al. examined the changes in serum N-linked glycans of a patient with sepsis during the first eight days⁷⁶. The relative amount of sialyl Lewis glycans to core fucosylated glycans rose from 34% in a control sample to 52% and 63% on Day 1 and Day 8, respectively, in the sepsis samples. These results imply that the proportion of different fucosylated glycan structures changes daily, which apparently mirrors the disease since N-glycan profiles in the healthy serum remained approximately constant. Despite extensive studies focusing on changes in fucosylated glycans, for other reliable biomarkers of disease, only AFP is currently

used for serologic monitoring of HCC. Even this biomarker still has relatively high false-negative rates in HCC categorization⁸⁶.

III. N-acetylglucosamine (GlcNAc) branches on newly synthesized glycoproteins, initiated by the medial Golgi N-acetylglucosaminyltransferases I, II, IV, and V (Mgat1, Mgat2, Mgat4a/b/c and Mgat5), is thought to be linked to tumor progression^{87,88}. The Mgat5 enzyme is responsible for adding β -1,6-GlcNAc to N-glycans, producing the most complex types of N-glycans, such as tetra-antennary and poly N-acetylactosamine (LAcNAc) structures. Granovsky et al. have shown that Mgat5-deficient mice with carcinomas display a longer latency and fewer metastases⁸⁹. Another enzymatic product of branching N-glycan is the addition of the bisecting GlcNAc to mannose in β -1,4-linkage catalyzed by the Mgat3. Recently, Tan et al. showed that bisecting GlcNAc structures hampers the epithelial-mesenchymal transition (EMT) process, which commonly plays a key role in cancer progression, suggesting the therapeutic potential of enhancing Mgat3 as a treatment for inhibiting breast cancer development⁹⁰. Sarrats et al. reported that changes in N-glycan branching on some acute-phase proteins also influence the response to inflammation in both chronic pancreatitis (CP) and pancreatic cancer (PAC) patients⁶². They found that the amount of tetra-antennary structures increased on the α -1 acid glycoprotein (AGP) and haptoglobin (HPT). An increase in the amount of bisected and tri-antennary N-glycans was also observed on transferrin (TRF) in both patients. Although knowledge of the synthesis of N-glycan branches in cancer cells has improved, there are still obstacles to its clinical application. Technical advances on the sensitivity and specificity of the targeted N-glycan branches need to be improved.

To summarize, it has become increasingly evident that changes in glycosylation can be considered as hallmarks of several steps in human disease pathogenesis. Serious efforts have already been made revealing relationships between changes in N-glycosylation on specific proteins and various diseases. However, most of these studies are based solely on a small number of tumor samples or serum and thus still lack clinical validation. Further studies are needed to use carefully selected tumor samples and larger cohorts and or longitudinal sampling to uncover the role of protein glycosylation changes in every step of the tumor progression or in determining the specific disease states.

2. Mass spectrometry-based methods for Characterizing Protein N-glycosylation

The analysis of glycoproteins by mass spectrometry is a quickly expanding field. Glycoproteins directly or indirectly influence many molecular interactions in cells and body fluids. Therefore, characterizing the structure and function of glycoproteins is crucial for unraveling the glycosylation “code” and developing glycoprotein-based therapeutics⁹¹. However, characterizing protein N-glycosylation has specific concerns that make the analysis challenging. One of the problems is that the process of glycan biosynthesis on the protein is not controlled by a template. Thus, there is a lack of a biological amplification method (as for DNA), which would provide ample amounts of glycoproteins. As a result, glycoproteins usually have to be analyzed at their physiological concentrations. Low amounts of sample require development of sensitive analytical methods that can detect low abundant molecules in complex matrices. A second major challenge is the inherent complexity of glycan structures resulting in the macro- and microheterogeneity of glycosylation, thereby forming various glycoforms. One glycosylation site can exhibit more than a dozen different glycoforms; however, examples of even up to 100 glycoforms have been reported⁹². Potentially, each glycoform may exhibit a different biological function, and addressing them will represent a long-lasting challenge. Over the past two decades, various analytical strategies for characterizing protein glycosylation have been pioneered^{93,94,95,96}. Currently, analyzing protein N-glycosylation by mass spectrometry can roughly be divided, based on the

molecular entities investigated, into three different approaches: analyzing the glycans released from glycoproteins (glycomics approach), analyzing intact glycopeptides (glycoproteomics approach), and analyzing intact glycoproteins (denaturing and native MS approach).

Through the glycomics approach, individual glycan structures are analyzed in detail, including the monosaccharide composition, and the sequence, branching, and linkage. This approach has been extensively used in early MS analyses, for instance, in glycan-based biomarker discovery studies for diagnosis of early-stage human cancers, monitoring disease progression and response to therapy, and detecting recurrence^{97,98}. Such studies typically monitor the levels of fucosylation or sialylation in the detected N-glycans and compare samples derived from various physiological states, for example, diseased vs. healthy people⁹⁹. Since the glycomics approach does not preserve information about site-specificity, it is nowadays more and more combined with also analyzing the intact glycopeptides¹⁰⁰.

The peptide-centric glycoproteomics approach can provide a site-specific characterization of N-glycosylation, including site-occupancy and microheterogeneity. Glycoproteomics technologies currently focus on advancing the development of biomolecular mass spectrometry, including improved fragmentation methods and bioinformatics tools for mapping disease-associated aberrancies in the glycoproteome. These aberrancies originate from different tissues, cells, and body fluids, which may be regulated differently during pathogenesis⁶⁹. In general, profiling intact glycopeptides enriched from cells or body fluids may reveal the glycosylation (fucosylation, sialylation, and branching) status of proteins and provide detailed information for the specific alternation in glycan composition comparing people with healthy and disease states. Combining both glycomics and glycoproteomics approaches can determine which sites on each glycoprotein are glycosylated and where precisely these disease-specific alternations occur. Such information can be used for better-classifying disease or tracking disease progression¹⁰¹.

The third major, but so far least developed, approach for glycoprotein analysis is the intact MS approach, which can be further divided into denaturing intact MS (often also termed Top-Down MS) and native MS. Top-down MS¹⁰² has been used extensively for the analysis of other site-specific PTMs on proteins, for example, in the analysis of protein phosphorylation, acetylation, and methylation^{103,104}. Characterization of glycoproteins by Top-down MS is not yet that well established, although some promising studies have been recently reported⁶⁸. On the other hand, in particular, by work from also our laboratory, native MS has already shown great prominence in capturing and identifying proteoforms of glycoproteins^{105,106}. By measuring glycoproteins in their physiologically more relevant native state (meaning preserving its native conformation, including tertiary and quaternary structures), it is possible to reveal information not only about glycosylation status, but also about the site occupancy, protein-ligand binding, and protein-protein interactions^{106,107,108}. In the following paragraphs, I will briefly describe some of the current analytical strategies used in glycomics, glycoproteomics with advanced LC-MS/MS fragmentation methods, followed by a description of the native MS approach for the analysis of intact glycoproteins and finally I describe a hybrid MS approach for the more detailed analysis of intact glycoprotein that I developed and used for much of the work described in my thesis.

2.1 Mass spectrometry-based methods for Glycomics Analysis

The glycomics methods are used to characterize glycans released from either a selected set or all glycoproteins, originating from different cells, tissues, or even whole organisms. MS-based profiling of glycans has become a relatively high-throughput method for screening complex glycan mixtures and offers a comprehensive view on glycan repertoire in a given sample¹⁰⁹. Moreover, MS/MS methods such as collision-induced dissociation (CID) are suitable for producing glycan

fragment ions that allow the elucidation of the glycan composition and linkages between the individual glycan residues¹¹⁰. Therefore, glycan profiling has been frequently used for monitoring aberrant glycosylation in diseases, and even applied to relatively large cohort studies^{111–113}.

Essential techniques for glycomics experiments mainly include strategies for the enzymatic release of glycans and glycan derivatization. The most common and effective method for releasing N-glycans is the enzymatic removal using the peptide-N-glycosidase F (PNGase F) after digesting glycoprotein into peptides. PNGase F is an amidase, which cleaves the glycosidic bond between the innermost GlcNAc and asparagine (Asn) residues. PNGase F can efficiently release all kinds of high-mannose, complex, and hybrid N-glycans, except the ones containing α 1,3-linked core fucose moieties¹¹⁴. Alternatively, a similar enzyme, PNGase A from almonds, cleaves N-glycans with/without core fucose residues, but the efficiency of PNGase A to release glycans is somewhat lower than that of PNGase F. During the deglycosylation reaction, all formerly-glycosylated Asn are converted into aspartic acid (Asp) residues through deamidation, inducing an increase in peptide mass of 0.98 Da. A larger, more easily detectable, mass shift can be achieved by ¹⁸O-labeling of the resulting Asp using isotopically labeled water, which facilitates the peptide assignment and thus identification of the glycosylation site in the database search¹¹⁵. Unlike PNGase F and A, endoglycosidases specifically release different types of N-glycans. For example, endoglycosidase H (Endo H) is usually used to exclusively release high mannose and hybrid types of N-glycans between the innermost GlcNAc residues, leaving a GlcNAc attached to the Asn and adding a mass of 203.08 Da to the mass of the peptide backbone. The sequential treatment of enzymes with different specificity provides a useful strategy for characterizing complicated glycoproteins with multiple N-glycosylation sites. For example, Liwei et al. used this strategy to characterize the HIV-1 envelope glycoprotein (Env) containing 26-30 N-linked glycosylation sites with multiple types of glycan structures occupying each of these sites¹¹⁶. After the Env protein was first digested into peptides, the authors first used Endo H to remove all high mannose and hybrid glycans. The remaining peptide mixtures were subsequently treated, in the presence of H₂¹⁸O, with PNGase F, removing the complex glycans. As a result, each step of enzymatic treatment introduced unique mass-shifts to the glycopeptides, which was used to infer glycosylation sites that carried no glycans, high mannose/hybrid, and complex types of glycans.

A very fruitful strategy to tackle certain challenges related to glycan analysis is by using specific chemical modifications. Glycans are typically tricky to ionize, important for mass spectrometry, due to their hydrophilic characteristics and lack of basic residues amenable to become protonated during the electrospray process¹¹⁷. During the ionization process, the high hydrophobicity of analytes can ensure efficient droplet desolvation¹¹⁸. In addition to the poor ionization efficiency of glycans, the larger glycans are particularly difficult to analyze as they often contain labile fucose and sialic acid moieties, which are either lost during sample preparation or MS analysis before the ions reach the detector¹¹⁹. Therefore, glycan derivatization has become a common strategy to improve the stability and ionization efficiency of glycans during MS analysis. Reductive amination is one of the most used glycan labeling strategies for derivatizing glycans at their reducing end. This modification improves the ionization efficiency and stability of glycans with labile groups, such as sialic acid, allowing for a more reliable glycomic analysis. Methods whereby glycans are labeled with 2-aminobenzamide (2-AB) or 2-aminobenzoic (2-AA), for example, have been developed and widely used to enhance the analysis of glycans in positive ion MS mode⁹⁸. These labels can be attached to a chromophore or fluorophore to directly monitor and quantify the abundance of glycans using fluorescence¹²⁰ or UV detection¹²¹ in HPLC approaches. Permethylolation is another glycan derivatization approach that can convert negatively charged carboxylate residues on the glycans to neutral methyl esters, thereby improving the ionization efficiency, especially useful for sialylated glycans when using positive-ion MS mode¹²². Recent advancements in glycan derivatization enabled the differentiation of the α -2,3 and α -2,6 linkages

of sialic acid residues. Reiding et al. used therefore 1-ethyl-3-(3-(dimethylamino)propyl)-carbodiimide (EDC) and 1-hydroxybenzotriazole (HOBt) as activators to ethyl esterify α -2,6-linked sialic acids and lactonization of α -2,3-linked sialic acids variants to analyze released N-glycans. Using this approach, they described more than 100 distinct N-glycan compositions with varying sialic acid linkages present in human plasma¹²³.

2.2 Methods for Glycopeptide Enrichment and Separation

The standard glycoproteomics experiment involves digestion of protein(s) extracted or purified from various biological materials, which generates complex peptide mixtures containing non-modified and peptides modified by a variety of PTMs. Most of the glycopeptides in these peptide mixtures are low abundant compared to the non-modified peptides, and therefore it is quite challenging to detect them with MS directly. Moreover, glycopeptides are expected to have relatively lower ionization efficiencies, and their MS signal is distributed over many different glycoforms, which further complicates their analysis^{124,125}. Therefore, it is often essential to efficiently enrich and pre-fractionate glycopeptides prior to MS analysis. There are quite a few distinct enrichment and separation strategies. These include several lectin-based enrichment methods, hydrophilic interaction liquid chromatography (HILIC), solid-phase extraction methods (SPE), azido sugar metabolic labeling approaches, capillary electrophoresis (CE) separation, and ion mobility spectrometry (IMS) separation. More generic peptide separation approaches, i.e., also used for unmodified peptide analysis, such as in-gel separation and reverse-phase chromatography (RPLC), are also inevitable parts of peptide-centric glycosylation analysis, but those will not be further discussed here.

Lectin-based enrichment methods have been widely used to enrich both glycoproteins and glycopeptides carrying distinct types of N-glycans. Several lectins, such as Aleuria aurantia lectin (AAL, specific for fucosylated glycans), concanavalin A (Con A, specific for mannose or glucose-containing glycans) and wheat germ agglutinin (WGA, specific for glucose or sialic acid-containing glycans), have been extensively used to characterize and quantify various glycoforms in complex samples¹²⁶. Due to the ability of lectins to selectively capture specific carbohydrate moieties, recent clinical studies used a multi-lectin strategy to enrich specific types of glycans for more targeted analysis. Sarah et al. combined AAL and Phaseolus vulgaris erythroagglutinin (PHA-L/E, specific for branched glycans) to enrich and separate core-fucosylated and highly-branched protein glycoforms in the serum of prostate cancer patients and used this approach to analyze changes in glycosylation¹⁸. The lectins used in this study were chosen based on the aberrant changes of core-fucosylated and highly-branched glycans reported in patients with prostate cancer.

Multi-lectin approaches can enrich many different types of glycans in complex biological samples. However, they are inherently hampered by a bias towards specific glycans. Moreover, the co-occurrence of fucosylated and branched N-glycans on a protein may influence the PHA-L lectin binding⁹⁵. For example, Di Wu et al. reported that the PHA-L lectin recognizes and binds tri- and tetra-antennary N-glycans with multiple fucoses on α -1-acid glycoprotein; however, the same fucosylated N-glycans with two and three antennas on haptoglobin attenuated lectin binding⁹⁵. This data indicates that multi-fucosylation decreased PHA-L interactions with less branched N-glycans. Thus although powerful, lectin-based enrichment can be heavily biased towards the enrichment of selected parts of the glycome.

An alternative approach to the lectin-based approaches, which potentially is less biased, uses the unique chromatographic properties of glycopeptides. A method that has become very popular for glycopeptide enrichment is HILIC, available through a variety of distinct resins such as

cotton¹²⁷, sepharose¹²⁸, and zwitterionic stationary phases (ZIC-HILIC)⁹³. HILIC separation depends largely on the hydrophilicity and size of the analytes, which allows intact glycopeptides with highly hydrophilic glycans to be nicely separated from the more abundant smaller and less hydrophilic non-glycopeptides. However, HILIC affinity to the glycopeptide becomes weaker when the glycan on the peptide is becoming too long, increasing the risk of losing the glycopeptide during enrichment¹²⁹. Therefore, a combination of different enrichment methods has been introduced to improve glycopeptide coverage further. HILIC can be relatively easily coupled with other enrichment methods, enabling two-dimensional separation and fractionation. For example, Totten et al. combined electrostatic repulsion HILIC with strong anion exchange (SAX) solid-phase extraction, which significantly boosted the number of identified N-linked glycopeptides from human plasma, especially compared to lectin-based and HILIC-based methods¹⁹. Over 800 unique glycopeptides from 95 plasma glycoproteins were reported, making electrostatic repulsion HILIC coupled with SAX a very effective strategy for system-wide and site-specific mapping of the glycoproteome.

Another strategy for glycopeptide enrichment integrates chemical and enzymatical reactions. Sun et al. designed an up-and-coming method combining solid-phase extraction of N-linked glycans with glycosite-containing peptides (NGAG)¹³⁰. In this chemoenzymatic method, peptides obtained by glycoprotein digestion were covalently conjugated to aldehyde-functionalized beads followed by sequential chemical and enzymatic reactions to release the glycans. Subsequently, glycosite-containing peptides were released from the beads. This NGAG approach allowed for the large-scale profiling of glycans, glycosylation sites, and glycoproteins, all combined in a single analysis.

Another strategy developed by the Bertozzi lab combines the metabolic labeling of the glycoproteome with chemical enrichment and isotopic recoding of glycopeptides enabling the proteome-wide identification of both intact N- and O-glycopeptides^{131,132}. In this approach, glycoproteins are tagged by a multifunctional probe using click chemistry and enriched using streptavidin-agarose affinity chromatography. Retained glycoproteins are on-bead digested, and the generated isotopically recoded glycopeptides are eluted and analyzed by LC-MS/MS. The stable isotope tag is utilized in the pattern-searching algorithm to direct MS/MS analysis to only isotopically labeled species, which boosts the confidence of the glycopeptide identification. This targeted glycopeptide method significantly enhances the number of identified glycopeptides in a whole-cell glycoproteome approach.

The unique physical-chemical properties of glycoproteins can also be utilized for their separation based on electrophoretic mobility. Currently, glycoproteins are often separated using capillary electrophoresis (CE) coupled with MS (CE-MS). In this way, the sample heterogeneity may be reduced, and a detailed analysis of many proteins such as recombinant erythropoietin¹³², haptoglobin, and IgG¹³³ has become feasible. CE separation of glycopeptides and glycoproteins occurs in the liquid phase and depends on the charge, size, and shape. For example, CE can excel in separating glycopeptides/glycoproteins with the same backbone sequence, but with a different number of negatively charged sialic acids¹³³. However, CE approaches also present some inherent difficulties, such as the requirement of working with nL sample volumes or problems with establishing at the end of the CE capillary a stable electrospray ionization source, which have hampered the wider application of CE-MS for large-scale glycoproteomics so far. As in CE separations, similar physical-chemical properties of glycoproteins are employed in the analysis by IMS coupled to MS (IMS-MS). Recently, IMS-MS has gained traction in many areas of proteomics, including glycoproteomics^{134,135}. IMS separation is based on the collision between neutral buffer gas molecules and analyte ions in the gas phase, whereby the mobility is affected by the shape and charge of the ions¹³⁶. This mechanism allows for distinguishing glycopeptide ions with identical mass but different glycan structures. For example, peptides attaching either a

GlcNAc or GalNAc (isobaric isomers) can be separated and detected by IMS-MS¹³⁷. Advances in IMS-MS brought several applications for the detailed structural and site-specific analysis of glycopeptides^{138,139}. However, software tools for analyzing IMS-MS glycoproteomics data are currently far less mature than those available for proteomics and needs to be further developed.

To conclude, improvements in MS-based glycoproteomics analyses enable us to simultaneously obtain glycosylation sites and associated glycoforms for exploring glycan functions, not only in different body fluids but also on the surface of various cells. Although a common caveat of any glycoproteomic experiment is that there is no universal or ideal enrichment method¹⁹, ZIC-HILIC, with its features as convenience and speed, is still one of the most used methods for high-throughput glycoproteomic studies.

2.3 Tandem MS Methods targeting glycopeptides

Mass spectrometers cannot only measure the mass of intact glycopeptide ions (MS1 level) but also fragment ions thereof (MS2 level), produced by a variety of tandem MS methods (MS/MS). Acquired MS/MS spectra are used in a database search, providing identification of the glycoprotein they originated from, the peptide sequence information, the glycan compositions, and the site localization of the PTMs. Here I will describe in brief the different tandem MS methods that have been applied most frequently to characterize intact glycopeptides.

Low energy collision-induced dissociation (CID) is still the most commonly used MS fragmentation technique. In CID, precursor ions collide with gas molecules (helium, nitrogen, or argon) multiple times in the gas phase. These collisions increase the internal energy of the ions sufficiently to break the (weaker) amide bonds (C-N bonds) of the peptide backbone, resulting in a series of b (N-terminal fragments) and y fragment ions (C-terminal fragments)¹⁴⁰. However, during the collision of intact glycopeptides, the glycan moieties compete with the amide bonds of the peptide backbone for the preferred site of protonation and thus also preferred sites of fragmentation. These fragmentation events lead to cleavage of glycosidic linkages of glycopeptides and production of glycan ion series, which are called B (glycan attached to N-terminal fragment) and Y (glycan attached to C-terminal fragment) fragment ions¹⁴¹. Although B and Y ions can provide structural information on the glycans, the lack of peptide fragment ions may cause difficulties in identifying the peptide sequence of the fragmented glycopeptides. A related fragmentation technique termed higher-energy collisional dissociation (HCD) also produces b and y fragment ions being, in principle, very similar to CID. Nevertheless, compared to CID, HCD increases the efficiency of dissociating the glycopeptide backbone while preserving the attached glycan¹⁴². HCD is mostly performed in Orbitrap-based mass analyzers, while CID is normally performed in ion traps, which is often limited by the “one-third rule”¹⁴³, resulting in the low trapping efficiency of the low m/z range ions. Although HCD can produce more peptide fragment ions than CID, B and Y are still in abundance predominantly observed ions in HCD MS/MS spectra due to the facile neutral loss of the glycan moiety¹⁴⁴. The neutral loss issues occurring for labile PTMs during both CID and HCD fragmentation initiated the development of alternative types of fragmentation methods. Electron-based methods, electron-capture dissociation (ECD)¹⁴⁵, and electron transfer dissociation (ETD)¹⁴⁶, have been introduced and brought about major advancements in the MS field in the last years. Although quite similar, there are some differences between the fragmentation mechanism of ECD and ETD. ECD proceeds via the direct capture of low energy electrons by multiply protonated ions¹⁴⁵. ETD proceeds via ion-ion reactions between radical anions (as electron donor) and peptide cation inside an ion trap. These ion-ion reactions lead to the transfer of electron to the peptide cation, forming a peptide ion radical. In both cases, the electron capture/transfer leads to peptide cation radicals with much internal energy making them somewhat unstable. They, therefore, quickly fragment, primarily

through the scission of N-C α bonds. These fragmentations of the peptide backbone produce distinctive fragment ions, notably mostly c (N-terminal fragment) and z (C-terminal fragment) ions. Specific targeting the peptide backbone is a significant advantage of ECD/ETD over CID/HCD, especially for glycopeptides, since it reduces glycosidic bond cleavage and retains intact glycan moieties on the peptide. Therefore, ECD/ETD spectra contain often more precise information on the glycosylation sites of glycopeptides. Although both ECD and ETD have been successfully applied to glycopeptide characterization^{142,147,148}, they are still not the most prominent approaches for intact glycopeptide characterization. In the case of ECD, it could till recently only be used on high-cost Fourier Transform-Ion Cyclotron Resonance (FT-ICR) MS¹⁴⁹, albeit its capabilities have recently evaluated by incorporating ECD into a benchtop Orbitrap Q Exactive MS¹⁴². In contrast, ETD is available in somewhat lower-cost mass analyzers such as ion trap or Orbitrap. A major limitation of these electron induced fragmentations lies in their relative poor dissociation efficiency, especially for low charged precursors, notably also glycopeptides¹⁴³. Overall, all commonly used fragmentation methods for intact glycopeptide analysis, including CID, HCD, ECD, and ETD have their pros and cons. Therefore, combining different fragmentation mechanisms seems to be an ideal step forward in achieving a better analysis of intact glycopeptides. These so-called hybrid fragmentation methods can be divided into 1) combined different glycopeptide fragmentation methods, such as HCD/CID with ETD (EThcD, ETciD)¹⁵⁰, 2) product ion-dependent triggered methods such as HCD product ion-dependent EThcD (HCD-pd-EThcD)^{151,152}, and 3) combined different HCD collision energies (stepped HCD)¹⁵³.

EThcD is a cutting-edge hybrid fragmentation method, integrating HCD and ETD, which generates MS/MS spectra containing fragment ions induced by both methods. This approach has been demonstrated as a significant advancement in glycopeptide identification¹⁵⁴. Due to low ETD efficiency, ETD MS/MS spectra contain unfragmented and charge-reduced precursor ions. The presence of this highly abundant precursor ion could be an inspiration to develop a two-fold fragmentation strategy where these precursor ions can be further subjected to collisional activation in the HCD cell. This supplemental energy for ETD reactions increases the efficiency of ETD fragmentation and provides in addition to c and z ions, b, and y type ions. One issue of applying EThcD is that its speed is relatively slower compared to especially when HCD fragmentation is used alone. Therefore, the mass spectrometer spends more time on fragmenting non-glycopeptides in the complex samples, which results in lower glycopeptide identification.

The great versatility of new hybrid mass spectrometers offers a solution to this issue, as these instruments allow not only combining different fragmentation mechanisms but also their targeted use during the MS experiment. As mentioned above, glycopeptide HCD MS/MS spectra provide many diagnostic oxonium ions, such as *m/z* 204.087 (HexNAc), 163.06 (Hex), 292.10 (Neu5Ac) and 366.14 (HexNAc-Hex), which may serve to confirm that the precursor ion is a glycopeptide. Moreover, these abundant oxonium ions can be utilized for using product-dependent triggered fragmentation methods. In such workflows, fast HCD or CID MS/MS are set as primary fragmentation events throughout the whole MS run, and slow hybrid MS/MS events are triggered only at the moment the specific oxonium ions are detected in the prior fast MS/MS event. Such smart algorithms provide better time-management helping the mass spectrometer to spent more analysis time on the glycopeptides of interest, thereby largely increasing the chance of detecting low abundant glycopeptides in complex peptide mixtures. In this manner, HCD-pd-EThcD method has become one of the most potent fragmentation strategies for large scale glycopeptide studies¹⁵².

Another example of a hybrid fragmentation approach is stepped HCD, where a precursor ion is fragmented by multiple collision energies (typically three different energies from low to high). Higher HCD collision energies provide better peptide backbone fragments (b and y ions), while

lower collision energies facilitate producing glycan fragments (B and Y ions)¹⁵⁵. All fragment ions acquired from different collision energies can be merged into one more informative MS/MS spectrum, which significantly improves the large scale glycoproteome analysis¹⁵⁶.

To summarize (Figure 4), the development of these fragmentation methods, especially for the stepped HCD and the HCD-pd-ETHcD have been widely used for the in-depth glycosylation sites mapping and intact glycopeptides identification in complicated biological samples¹⁵⁶.

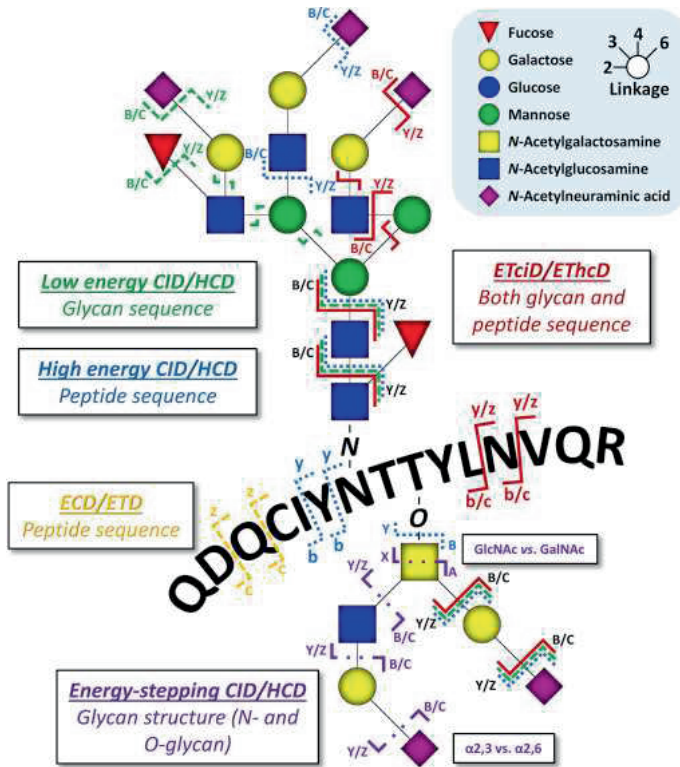


Figure 4. Overview of frequently used MS/MS fragmentation methods to analyze glycopeptides. B (N-terminal) and Y (C-terminal) types ions are glycopeptide ions, resulting from the sequential neutral losses of terminal monosaccharides. b,c (N-terminal) and y,z (C-terminal) ions are peptide backbone fragment ions. Figure adapted with permission from¹⁵⁷

2.4 Intact and Hybrid MS methods for Glycoprotein Analysis

Up to date, most of the glycoprotein characterization studies have been based on peptide-centric methods as described earlier, which are primarily targeted at the high-throughput analysis of complex protein mixtures, reaching the level of the entire cellular or tissue glycome or glycoproteome. These methods can also be used to analyze in-depth a single glycoprotein, providing a landscape of glycan composition, glycosylation site, and peptide sequence. Although peptide-centric MS approaches have been widely adopted and represent a powerful toolbox for

glycoprotein characterization, the analysis of glycoproteins at the peptide level also has some limitations. The analysis of glycoproteins by cutting them into small glycopeptides represents an obvious approximation, and some important information on the whole protein may be easily necessarily lost. For instance, glycosylation may be involved in intramolecular PTM interplay or depend on the co-existence of some other structural features, which cannot be preserved due to the protein-destructive nature of peptide-centric approaches. Another well-known problem in peptide-centric approaches is the observed facile loss of labile glycans such as sialic acids, which can lead to a bias in the determination of the protein sialylation levels. Furthermore, identifying various specific peptide modifications relies heavily on the parameters, which are manually set by the user in the search software, i.e., if a certain modification is not searched for, it will not be found. Therefore, unexpected modifications, sequence variants, or mutations can easily be omitted and left unnoticed.

Some of the limitations of peptide-centric approaches may be overcome by protein-centric methods such as top-down MS and native MS analysis. These methods directly analyze glycoproteins by MS, circumventing digestion. With protein-centric approaches, it is possible to capture and characterize proteoforms of glycoproteins and directly access PTMs in the context of their entire structure, including the stoichiometry of (glyco)protein complexes. Such analysis, applied to an intact glycoprotein, provides unique information giving an overview of the type, abundance, and extend of glycosylation. When the analysis is performed either under denaturing or native conditions, the acquired peak series in the MS spectra of the intact glycoproteins exhibit a coherent nature and can be considered as the complete information about the glycoprotein heterogeneity. Observed mass differences between detected peaks can be typically attributed to the proteoforms differing from each other by specific PTMs, amino acids, ligands, metal ions, or other modifications. Therefore, such intact glycoprotein data has a high discovery potential since full annotation of the intact MS spectra may reveal unknown or unexpected modifications, sequence variants, mutations, or other structural features that are more difficult or even impossible to be captured by peptide-centric approaches.

As mentioned before, the intact MS analysis of glycoproteins can be classified based on the conditions under which the protein samples are introduced into the mass spectrometer. Top-down MS analysis is usually performed in combination with LC or CE, which requires using MS-compatible buffers containing water, organic solvent, and acid to enhance the protein solubility, desolvation, and ionization during the ESI process. Although top-down MS analysis still needs substantial advancements, recent studies have demonstrated its ability to access the complexity of heterogeneous therapeutics^{158,159} and other complicated protein assemblies. For example, Nagel et al. applied a top-down MS approach to characterize sequence polymorphisms and glycosylation patterns of fibrinogen purified from human plasma¹⁶⁰. Moreover, top-down MS, coupled online with LC, has proven to be successful in the unambiguous identification of hundreds of proteins and proteoforms with a molecular weight below 30k Da extracted from cell lysates¹⁶¹. However, such analyses are still more complicated when considering the glycoproteome mainly due to the problematic data interpretation of glycoprotein mixtures, inefficient separations of glycoproteoforms, and limitations related to the MS measurements restricted to mass range within 5-50 kDa¹⁶².

Denaturing conditions used in top-down MS analysis inevitably disrupt protein higher-order structure, induce unfolding, loss of oligomeric states, and non-covalent interactions. Native MS analysis can prevent loss of this specific information as the proteins are analyzed under more physiological conditions. Therefore, native MS can simultaneously provide data on the function, stoichiometry, protein-ligand binding, and protein-protein interactions. One of the crucial requirements for native MS is to maintain proteins in their native-like conformations while they are being injected into the mass spectrometer via the ESI process. In most cases, ionizing the

proteins is performed from a volatile aqueous ammonium acetate buffer at physiological pH. This process allows analytes to be retained in their folded states and keeps some chargeable sites buried inside the protein structure. As a consequence, the ionized proteins take up fewer charges than under denaturing conditions, shifting the charge state envelop toward a higher m/z mass region. The reduced charge state of the analyte provides a larger space between adjacent charge states, which is particularly useful for characterizing glycoproteins with multiple glycosylation sites and heterogeneous glycans. As illustrated in **Figure 5a**, under denaturing conditions, all the signal ions contain a high number of charges centering them to a relatively low m/z region in a narrow m/z window. Unfortunately, these overlapping charge states are difficult or impossible to resolve. On the other hand, the native MS spectrum shows the same ion series in a higher m/z region due to their reduced charge states, which enables their separation and resolves mass differences induced by PTMs. Such high resolving power has become available with the advent of high-resolution mass spectrometers such as the Orbitrap EMR (extended mass range)¹⁶³ and UHMR (ultra-high mass resolution)^{164,165}. With these new technologies, native MS has started to be applied in the field of glycoprotein therapeutics. For example, Caval et al. demonstrated that high-resolution native MS could characterize dozens of glycoengineered variants of EPO in a fast, sensitive, and reproducible manner¹⁰⁵. Native MS also started to meet the demand for reliable and accurate qualitative and quantitative analysis of monoclonal IgG glycoproteoforms^{166,167,168}.

Like any other method, also native MS has drawbacks. One of the major concerns is to keep the structure of proteins in their native conformations throughout the entire process of sample handling, including purification, fractionation (if applied), and final steps conducted during the sample preparation protocol for native MS¹⁶⁹. Despite many improvements toward automating and making native MS more “user-friendly,” it remains somewhat a low-throughput method requiring a well-trained operator. Expert care is also required to handle data processing and interpretation. In this regard, academia and industry have invested much effort into the development of software for data analysis^{170,171,172}, albeit usage of these tools still requires manual inspection and human judgment. For example, relative quantification of proteoforms has to be done very carefully due to the inconsistent intensity distribution of signals corresponding to proteoforms among different charge states. (**Figure 5b**). This is likely due to ionization bias caused by the different ability of glycoforms to take on charges¹⁷³. Therefore, conventional deconvolution algorithms often have problems with these data and generate incorrect output, which can lead to misleading conclusions. Nevertheless, native MS is becoming a widely accepted method for product characterization in pharmaceutical companies, and further development will lead to the utilization of native MS not only for research but also for quality control of protein-based drugs and diagnostics.

Since both peptide- and protein-centric methods have their limitations in analyzing protein glycosylation, hybrid MS approaches combining different MS methods are gaining popularity for glycoprotein characterization (**Figure 6**). Compelling integrative workflow, combining peptide-centric MS and high-resolution native MS, has been introduced for comprehensive and unbiased characterization of single glycoproteins. The integration of these two different MS concepts enabled detailed site-specific characterization of glycoprotein proteoforms, including their full quantitative profile displayed in a single native spectrum¹⁷⁴. This approach has been applied for in-depth characterization of therapeutic proteins^{175,105} and protein complexes from various biological sources¹⁷⁶, including human serum^{177,178,179}. However, the challenge still lies in the complexity of glycoproteins having heterogeneous glycoproteoforms, containing both N- and O-linked glycans on multiple glycosites. Glycans with similar masses generate overlapping signal peaks of co-occurring glycoproteoforms in the native MS spectra. Therefore, intact proteins need to be usually treated with different glycoside hydrolases such as sialidase, galactosidase, or PNGase F, to reduce glycosylation heterogeneity. As an example, removing sialic acid reduces the complexity of the native MS spectra. This step is particularly essential to discern between

double fucosylation and sialylation since they have similar masses (i. e. 291.1 and 292.1 Da, respectively). In an illustrative study of Wohlschläger et al., the authors tackled the complexity of the fusion protein Etanercept (ETN) (Enbrel®) by combining native MS and a sophisticated strategy of enzymatic digestions¹⁷⁵. ETN consists of two copies of the tumor necrosis factor- α receptor (TNFR) and the Fc portion of human IgG1. This biopharmaceutical, which can be used to treat autoimmune diseases, contains four N-glycosylation sites and 26 O-glycosylation sites in its dimeric TNFR domain, and two N-glycosylation sites in the IgG Fc portion. First, the protein backbone was simplified by applying the protease IdeS cleaving ETN at the hinge region, which allowed analysis of the N-glycosylation of TNFR and Fc domains independently. Next, the glycosylation heterogeneity was reduced by using sialidase, PNGase F, and O-glycosidase. Eventually, the authors performed bottom-up MS analysis and integrated all data to achieve a full characterization of this extremely complex glycoprotein.

Since top-down MS can provide sequence information by conducting MS/MS, it is tempting to combine it with native MS in a similar fashion as was done for combined bottom-up and native MS approach. Although top-down MS has not yet been established for glycoprotein analysis, there are some recent studies in the literature, which demonstrated the potential of top-down MS to tackle the complexity of protein complexes¹⁰³ or even protein mixtures¹⁰⁴. One successful example of integrating top-down MS and native MS for analysis of glycoprotein haptoglobin demonstrates the current boundaries of this strategy⁸. In this work, the authors revealed remarkable integrity between the two advanced MS approaches, which contributed to discovering subtle differences in glycosylation patterns between different haptoglobin oligomers purified from blood serum of healthy individuals. Interestingly, the data showed that the composition of N-glycans on haptoglobin seems to be dependent on higher-order structure and haptoglobin genotype, which affect binding properties of haptoglobin to hemoglobin. Native MS analysis provided accurate Mw measurements of haptoglobin complexes covering broad m/z range from 5000 to 12,000, whereas top-down MS/MS coupled online with HILIC LC complemented the native MS data with detailed quantitative site-specific glycosylation profiles of all haptoglobin subunits.

To summarize, hybrid approaches for intact glycoprotein characterization, using different MS methods, chromatographic separation, ion mobility, and other state-of-the-art analytical methods touched upon in this thesis represent some future strategies and directions taken to characterize glycoproteins. Deep understanding of currently available MS concepts and knowledge of their capabilities drive the development toward their more efficient integration. Peptide-centric MS can identify glycan compositions and localize glycosylation sites on the glycoprotein. Top-down MS analysis can separate proteoforms and provide compositional information of glycans on the protein or protein-subunit level. Native MS appears to be very useful in revealing the overall glycoform profile of glycoproteins, including other PTMs and modifications. In the current state, bioinformatics tools to integrate and interpret data generated from different MS approaches need to be further developed, which will allow broader application of these methods. Also, it is important to improve protein separation methods suitable for MS coupling, which will resolve the sample heterogeneity and steer both native MS and top-down MS to proteome level. In conclusion, I am convinced that hybrid MS approaches offer a promising future for both large-scale and in-depth characterization of glycoprotein and glycoproteome, as I also hope to demonstrate in the remainder of this thesis.

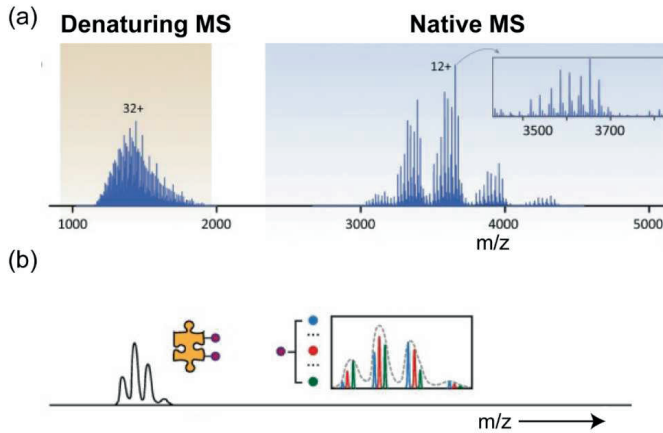


Figure 5. Structural heterogeneity of a glycoprotein measured by mass spectrometry. (a) High heterogeneity in protein glycosylation results in an extremely congested mass spectrum under denaturing conditions, whereas the signal peaks can be better separated and mass resolved in the mass spectrum under native condition. (b) The puzzle represents a glycoprotein with two glycosylation sites, which are indicated by pinheads colored in purple. Each glycosylation site may contain various proteoforms, of which a few are indicated by the three pinheads colored in blue, red, and green. The inset shows a close-up view of the ion signals of these three color-coded proteoforms, which have different charge state distributions. Such proteoform profiles may lead to misassignment of charges when deconvoluting the spectrum. Figure is adapted with permission from^{180,173}.

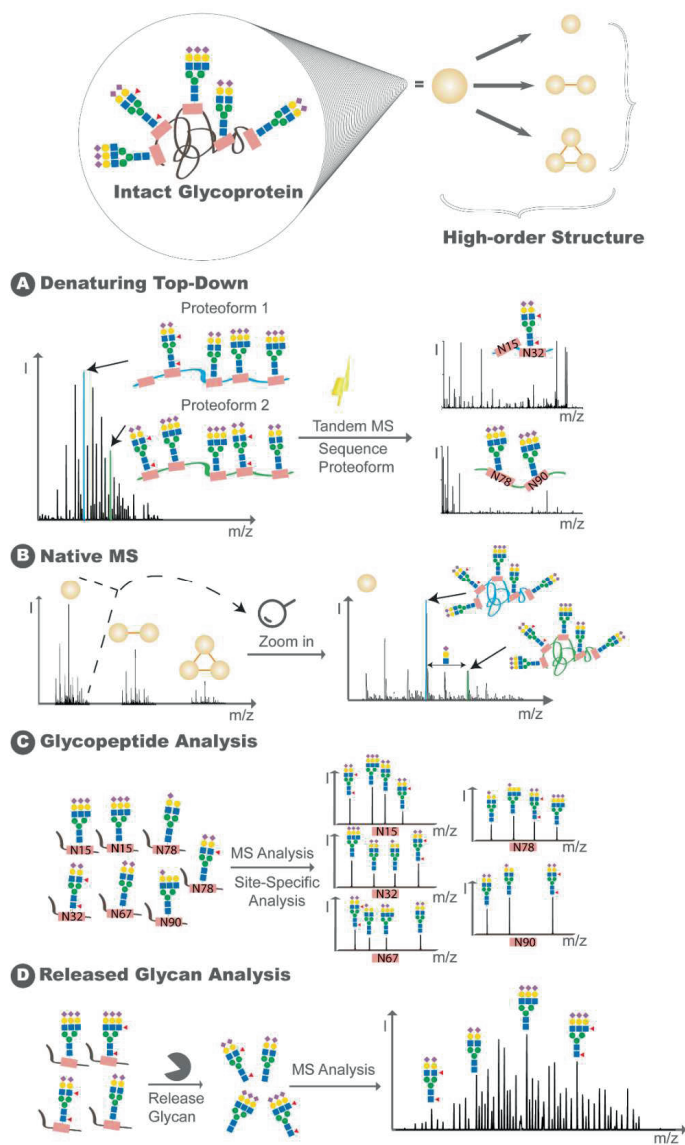


Figure 6. Overview of hybrid mass spectrometry (MS) based approaches for the characterization of glycoproteins. On top of the figure is a typical structure of a glycoprotein, which contains five N-glycosylation sites. Moreover, this glycoprotein potentially forms oligomeric higher-order assemblies (monomer and heteromer). (A) Top-down MS approach under denaturing condition can be applied to sequence the different proteoforms, with the aim to preserve the site-specific information of the intact glycans. (B) Native MS approach under non-denaturing conditions allows the investigation of the protein higher-order structure and provides information on the overall glycan combinations for each proteoform. (C) Glycopeptide analysis can be used to study site-specific glycosylation patterns quantitatively. (D) The released glycan approach can provide detailed structural information of the N-glycans.

Reference

- (1) Khoury, G. A.; Baliban, R. C.; Floudas, C. A. Proteome-Wide Post-Translational Modification Statistics: Frequency Analysis and Curation of the Swiss-Prot Database. *Sci. Rep.* 2011, 1, 90.
- (2) Belsole, R. J.; Greene, T. L. On the Frequency of Protein Glycosylation, as Deduced from Analysis of the SWISS-PROT Database. *Biochim Biophys Acta.* 1986, 1474 (1), 4-8.
- (3) Li, F.; Zhang, Y.; Purcell, A. W.; Webb, G. I.; Chou, K. C.; Lithgow, T.; Li, C.; Song, J. Positive-Unlabelled Learning of Glycosylation Sites in the Human Proteome. *BMC Bioinformatics.* 2019, 20, 112.
- (4) Leney, A. C.; El Atmioui, D.; Wu, W.; Ovaa, H.; Heck, A. J. R. Elucidating Crosstalk Mechanisms between Phosphorylation and O-GlcNAcylation. *Proc. Natl. Acad. Sci. U. S. A.* 2017, 114 (35), E7255-E7261.
- (5) Reily, C.; Stewart, T. J.; Renfrow, M. B.; Novak, J. Glycosylation in Health and Disease. *Nat. Rev. Nephrol.* 2019, 15 (6), 346-366.
- (6) Hofsteenge, J.; Müller, D. R.; de Beer, T.; Löffler, A.; Richter, W. J.; Vliegthart, J. F. G. New Type of Linkage between a Carbohydrate and a Protein: C-Glycosylation of a Specific Tryptophan Residue in Human RNase U_s. *Biochemistry.* 1994, 33 (46), 13524-13530.
- (7) Stepper, J.; Shastri, S.; Loo, T. S.; Preston, J. C.; Novak, P.; Man, P.; Moore, C. H.; Havlíček, V.; Patchett, M. L.; Norris, G. E. Cysteine S-Glycosylation, a New Post-Translational Modification Found in Glycopeptide Bacteriocins. *FEBS Lett.* 2011, 585 (4), 645-650.
- (8) Varki, A. Biological Roles of Glycans. *Glycobiology.* 2016, 27 (1), 3-49.
- (9) Ohtsubo, K.; Marth, J. D. Glycosylation in Cellular Mechanisms of Health and Disease. *Cell.* 2006, 126 (5), 855-867.
- (10) Chung, C. H.; Mirakhur, B.; Chan, E.; Le, Q. T.; Berlin, J.; Morse, M.; Murphy, B. A.; Satinover, S. M.; Hosen, J.; Mauro, D.; et al. Cetuximab-Induced Anaphylaxis and IgE Specific for Galactose- α -1,3- Galactose. *N. Engl. J. Med.* 2008, 358 (11), 1109-1117.
- (11) Qian, J.; Liu, T.; Yang, L.; Daus, A.; Crowley, R.; Zhou, Q. Structural Characterization of N-Linked Oligosaccharides on Monoclonal Antibody Cetuximab by the Combination of Orthogonal Matrix-Assisted Laser Desorption/Ionization Hybrid Quadrupole-Quadrupole Time-Of-Flight Tandem Mass Spectrometry and Sequential Enzy. *Anal. Biochem.* 2007, 364 (1), 8-18.
- (12) Lalonde, M. E.; Durocher, Y. Therapeutic Glycoprotein Production in Mammalian Cells. *J. Biotechnol.* 2017, 251, 128-140.
- (13) Dumont, J.; Euwart, D.; Mei, B.; Estes, S.; Kshirsagar, R. Human Cell Lines for Biopharmaceutical Manufacturing: History, Status, and Future Perspectives. *Crit. Rev. Biotechnol.* 2016, 36 (6), 1110-1122.
- (14) Kim, J. Y.; Kim, Y. G.; Lee, G. M. CHO Cells in Biotechnology for Production of Recombinant Proteins: Current State and Further Potential. *Appl. Microbiol. Biotechnol.* 2012, 93 (3), 917-930.
- (15) Lai, T.; Yang, Y.; Ng, S. K. Advances in Mammalian Cell Line Development Technologies for Recombinant Protein Production. *Pharmaceuticals.* 2013, 6 (5), 579-603.

- (16) Steven M. Banik, Kayvon Pedram, Simon Wisnovsky, Nicholas M. Riley, C. R.; Bertozzi. Lysosome Targeting Chimeras for the Degradation of Secreted and Membrane Proteins. *ChemRxiv*. 2019.
- (17) Yang, Z.; Hancock, W. S.; Chew, T. R.; Bonilla, L. A Study of Glycoproteins in Human Serum and Plasma Reference Standards (HUPO) Using Multilectin Affinity Chromatography Coupled with RPLC-MS/MS. *Proteomics*. 2005, 5 (13), 3353-3366.
- (18) Totten, S. M.; Adusumil, R.; Kullolli, M.; Tanimoto, C.; Brooks, J. D.; Mallick, P.; Pitteri, S. J. Multi-Lectin Affinity Chromatography and Quantitative Proteomic Analysis Reveal Differential Glycoform Levels between Prostate Cancer and Benign Prostatic Hyperplasia Sera. *Sci Rep*. 2018, 8 (1), 1-13.
- (19) Totten, S. M.; Feasley, C. L.; Bermudez, A.; Pitteri, S. J. Parallel Comparison of N - Linked Glycopeptide Enrichment Techniques Reveals Extensive Glycoproteomic Analysis of Plasma Enabled by SAX-ERLIC. *J Proteome Res*. 2017, 16 (3), 1249-1260.
- (20) Lin, Y. H.; Zhu, J.; Meijer, S.; Franc, V.; Heck, A. J. R. Glycoproteogenomics: A Frequent Gene Polymorphism Affects the Glycosylation Pattern of the Human Serum Fetuin/ α -2-HS-Glycoprotein. *Mol. Cell. Proteomics*. 2019, 18 (8), 1479-1490.
- (21) Goettig, P. Effects of Glycosylation on the Enzymatic Activity and Mechanisms of Proteases. *Int J Mol Sci*. 2016, 17 (12), 1-24.
- (22) Zielinska, D. F.; Gnad, F.; Schropp, K.; Wiśniewski, J. R.; Mann, M. Mapping N-Glycosylation Sites across Seven Evolutionarily Distant Species Reveals a Divergent Substrate Proteome Despite a Common Core Machinery. *Mol. Cell*. 2012, 46 (4), 542-548.
- (23) Imberty, A.; Pérez, S. Stereochemistry of the N-Glycosylation Sites in Glycoproteins. *Protein Eng. Des. Sel*. 1995, 8 (7), 699-709.
- (24) Lam, P. V. N.; Goldman, R.; Karagiannis, K.; Narsule, T.; Simonyan, V.; Soika, V.; Mazumder, R. Structure-Based Comparative Analysis and Prediction of N-Linked Glycosylation Sites in Evolutionarily Distant Eukaryotes. *Genomics, Proteomics Bioinforma*. 2013, 11 (2), 96-104.
- (25) Stanley, P. Golgi Glycosylation. *Cold Spring Harb. Perspect. Biol*. 2011, 3 (4), 1-13.
- (26) Sethi, M. K.; Thaysen-Andersen, M.; Smith, J. T.; Baker, M. S.; Packer, N. H.; Hancock, W. S.; Fanayan, S. Comparative N-Glycan Profiling of Colorectal Cancer Cell Lines Reveals Unique Bisecting GlcNAc and α -2,3-Linked Sialic Acid Determinants Are Associated with Membrane Proteins of the More Metastatic/Aggressive Cell Lines. *J. Proteome Res*. 2014, 13 (1), 277-288.
- (27) Balog, C. I. A.; Stavenhagen, K.; Fung, W. L. J.; Koeleman, C. A.; McDonnell, L. A.; Verhoeven, A.; Mesker, W. E.; Tollenaar, R. A. E. M.; Deelder, A. M.; Wuhler, M. N-Glycosylation of Colorectal Cancer Tissues: A Liquid Chromatography and Mass Spectrometry-Based Investigation. *Mol. Cell. Proteomics*. 2012, 11 (9), 571-585.
- (28) Chatterjee, S.; Lee, L. Y.; Kawahara, R.; Abrahams, J. L.; Adamczyk, B.; Anugraham, M.; Ashwood, C.; Sumer-Bayraktar, Z.; Briggs, M. T.; Chik, J. H. L.; et al. Protein Paucimannosylation Is an Enriched N-Glycosylation Signature of Human Cancers. *Proteomics*. 2019, 19 (21-22), 1-13.

- (29) Kawano, S.; Hashimoto, K.; Miyama, T.; Goto, S.; Kanehisa, M. Prediction of Glycan Structures from Gene Expression Data Based on Glycosyltransferase Reactions. *Bioinformatics*. 2005, 21 (21), 3976–3982.
- (30) Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. The KEGG Resource for Deciphering the Genome. *Nucleic Acids Res*. 2004, 32 (90001), D277 - D280.
- (31) Nairn, A. V.; York, W. S.; Harris, K.; Hall, E. M.; Pierce, J. M.; Moremen, K. W. Regulation of Glycan Structures in Animal Tissues: Transcript Profiling of Glycan-Related Genes. *J Biol Chem*. 2008, 283 (25), 17298–17313.
- (32) Schachter, H. The Joys of HexNAc. The Synthesis and Function of N- and O-Glycan Branches. *Glycoconj. J*. 2000, 17 (7–9), 465–483.
- (33) Zacchi, L. F.; Schulz, B. L. N-Glycoprotein Macroheterogeneity: Biological Implications and Proteomic Characterization. *Glycoconj. J*. 2016, 33 (3), 359–376.
- (34) Fisher, P.; Spencer, H.; Thomas-Oates, J.; Wood, A. J.; Ungar, D. Modeling Glycan Processing Reveals Golgi-Enzyme Homeostasis upon Trafficking Defects and Cellular Differentiation. *Cell Rep*. 2019, 27 (4), 1231–1243.
- (35) Zolov, S. N.; Lupashin, V. V. Cog3p Depletion Blocks Vesicle-Mediated Golgi Retrograde Trafficking in HeLa Cells. *J. Cell Biol*. 2005, 168 (5), 747–759.
- (36) Toshihiko Oka, D. U.; Frederick M. Hughson, A.; Krieger, M. The COG and COPI Complexes Interact to Control the Abundance of GEARs, a Subset of Golgi Integral Membrane Proteins. *Mol. Biol. Cell*. 2004, 15, 2423–2435.
- (37) Papers, J. B. C.; Doi, M.; Miyagawa, S.; Murakami, H.; Takahagi, Y.; Nakai, R.; Yamada, M.; Murase, A.; Koyota, S.; Koma, M.; et al. Remodeling of the Major Pig Xenoantigen by N - Acetylglucosaminyltransferase III in Transgenic Pig *. *J Biol Chem*. 2001, 276 (42), 39310–39319.
- (38) Zhao, Y.; Sato, Y.; Isaji, T.; Fukuda, T.; Matsumoto, A.; Miyoshi, E. Branched N-Glycans Regulate the Biological Functions of Integrins and Cadherins. *FEBS J*. 2008, 275 (9), 1939–1948.
- (39) Re S, Miyashita N, Yamaguchi Y, Sugita Y. Structural Diversity and Changes in Conformational Equilibria of Biantennary Complex-Type N- Glycans in Water Revealed by Replica-Exchange Molecular Dynamics Simulation. *Biophys J*. 2011, 101 (10), L44–L46.
- (40) Dennis, J. W.; Nabi, I. R.; Demetriou, M. Metabolism, Cell Surface Organization, and Disease. *Cell*. 2009, 139 (7), 1229–1241.
- (41) Giovannone, N.; Smith, L. K.; Treanor, B.; Dimitroff, C. J. Galectin-Glycan Interactions as Regulators of B Cell Immunity. *Front. Immunol*. 2018, 9, 2839.
- (42) Norton, P.; Comunale, M. A.; Herrera, H.; Wang, M.; Houser, J. Development and Application of a Novel Recombinant Aleuria Aurantia Lectin with Enhanced Core Fucose Binding for Identification of Glycoprotein Biomarkers of Hepatocellular Carcinoma. *Proteomics*. 2016, 16 (24), 3126–3136.
- (43) Mondal, N.; Dykstra, B.; Lee, J.; Ashline, D. J.; Reinhold, V. N.; Rossi, D. J.; Sackstein, R. Distinct Human $\alpha(1,3)$ -Fucosyltransferases Drive Lewis-X/Sialyl Lewis-X Assembly in Human Cells. *J Biol Chem*. 2018, 293 (19), 7300–7314.

- (44) Gagneux, P.; Varki, A. Evolutionary Considerations in Relating Oligosaccharide Diversity to Biological Function. *Glycobiology*. 1999, 9 (8), 747–755.
- (45) Kizuka, Y.; Taniguchi, N. Enzymes for N-Glycan Branching and Their Genetic and Nongenetic Regulation in Cancer. *Biomolecules*. 2016, 6 (2), 25.
- (46) Schwenk, J. M.; Omenn, G. S.; Sun, Z.; Campbell, D. S.; Baker, M. S.; Overall, C. M.; Aebersold, R.; Moritz, R. L.; Deutsch, E. W. The Human Plasma Proteome Draft of 2017: Building on the Human Plasma PeptideAtlas from Mass Spectrometry and Complementary Assays. *J Proteome Res*. 2017, 16 (12), 4299–4310.
- (47) Anderson, N. L.; Anderson, N. G. The Human Plasma Proteome: History, Character, and Diagnostic Prospects. *Mol. Cell. Proteomics*. 2002, 1 (11), 845–867.
- (48) He, Y.; Ma, C.; Xing, J.; Wang, S.; Ji, C.; Han, Y. Serum Amyloid A Protein as a Potential Biomarker in Predicting Acute Onset and Association with In-Hospital Death in Acute Aortic Dissection. *BMC Cardiovasc Disord*. 2019, 19 (1), 282.
- (49) Geyer, P. E.; Wewer Albrechtsen, N. J.; Tyanova, S.; Grassl, N.; Iepsen, E. W.; Lundgren, J.; Madsbad, S.; Holst, J. J.; Torekov, S. S.; Mann, M. Proteomics Reveals the Effects of Sustained Weight Loss on the Human Plasma Proteome. *Mol. Syst. Biol*. 2016, 12 (12), 901.
- (50) Niu, L.; Geyer, P. E.; Wewer Albrechtsen, N. J.; Gluud, L. L.; Santos, A.; Doll, S.; Treit, P. V.; Holst, J. J.; Knop, F. K.; Vilsbøll, T.; et al. Plasma Proteome Profiling Discovers Novel Proteins Associated with Non-alcoholic Fatty Liver Disease. *Mol. Syst. Biol*. 2019, 15 (3), e8793.
- (51) Kim, M. S.; Pinto, S. M.; Getnet, D.; Nirujogi, R. S.; Manda, S. S.; Chaerkady, R.; Madugundu, A. K.; Kelkar, D. S.; Isserlin, R.; Jain, S.; et al. A Draft Map of the Human Proteome. *Nature*. 2014, 509 (7502), 575–581.
- (52) Chick, J. M.; Haynes, P. a; Bjellqvist, B.; Baker, M. S. Proteomics Research Articles. *J Proteome Res*. 2008, 7, 4974–4981.
- (53) Bellei, E.; Bergamini, S.; Monari, E.; Fantoni, L. I.; Cuoghi, A.; Ozben, T.; Tomasi, A. High-Abundance Proteins Depletion for Serum Proteomic Analysis: Concomitant Removal of Non-Targeted Proteins. *Amino Acids*. 2011, 40 (1), 145–156.
- (54) Nanjappa, V.; Thomas, J. K.; Marimuthu, A.; Muthusamy, B.; Radhakrishnan, A.; Sharma, R.; Ahmad Khan, A.; Balakrishnan, L.; Sahasrabudde, N. A.; Kumar, S.; et al. Plasma Proteome Database as a Resource for Proteomics Research: 2014 Update. *Nucleic Acids Res*. 2014, 42, 959–965.
- (55) Geyer, P. E.; Kulak, N. A.; Pichler, G.; Holdt, L. M.; Teupser, D.; Mann, M. Plasma Proteome Profiling to Assess Human Health and Disease. *Cell Syst*. 2016, 2 (3), 185–195.
- (56) Pernemalm M, Sandberg A, Zhu Y, et al. In-Depth Human Plasma Proteome Analysis Captures Tissue Proteins and Transfer of Protein Variants across the Placenta. *Elife*. 2019, 8, e41608.
- (57) Geyer, P. E.; Holdt, L. M.; Teupser, D.; Mann, M. Revisiting Biomarker Discovery by Plasma Proteomics. *Mol. Syst. Biol*. 2017, 13 (9), 942.
- (58) Geyer, P. E.; Voytik, E.; Treit, P. V.; et al. Plasma Proteome Profiling to Detect and Avoid Sample-related Biases in Biomarker Studies. *EMBO Mol. Med*. 2019, 11 (11), 1–22.
- (59) Polanski, M.; Anderson, N. L. A List of Candidate Cancer Biomarkers for Targeted Proteomics. *Biomark Insights*. 2007, 1, 1–48.

- (60) Jain S, Gautam V, Naseem S. Acute-Phase Proteins: As Diagnostic Tool. *J Pharm Bioallied Sci.* 2011, 3 (1), 118–127.
- (61) Kirwan, A.; Utratna, M.; Dwyer, M. E. O.; Joshi, L.; Kilcoyne, M. Glycosylation-Based Serum Biomarkers for Cancer Diagnostics and Prognostics. *Biomed Res Int.* 2015, 490531.
- (62) Sarrats, A.; Saldova, R.; Pla, E.; Fort, E.; Harvey, D. J.; Struwe, W. B.; de Llorens, R.; Rudd, P. M.; Peracaula, R. Glycosylation of Liver Acute-Phase Proteins in Pancreatic Cancer and Chronic Pancreatitis. *Proteomics - Clin. Appl.* 2010, 4 (4), 432–448.
- (63) Wang, M.; Long, R. E.; Comunale, M. A.; Junaidi, O.; Marrero, J.; Bisceglie, A. M. Di; Block, T. M.; Mehta, A. S. Novel Fucosylated Biomarkers for the Early Detection of Hepatocellular Carcinoma. *Cancer Epidemiol Biomarkers Prev.* 2009, 18 (6), 1914–1921.
- (64) Clerc, F.; Reiding, K. R.; Jansen, B. C.; Kammeijer, G. S. M.; Bondt, A.; Wuhrer, M. Human Plasma Protein N-Glycosylation. *Glycoconj. J.* 2016, 33 (3), 309–343.
- (65) Smith, E. R.; Ford, M. L.; Tomlinson, L. A.; Rocks, B. F.; Rajkumar, C.; Holt, S. G. Poor Agreement between Commercial ELISAs for Plasma Fetuin-A: An Effect of Protein Glycosylation? *Clin. Chim. Acta.* 2010, 411 (17-18), 1367–1370.
- (66) Langlois M. R.; Delanghe J. R. Biological and Clinical Significance of Haptoglobin Polymorphism in Humans. *Clin Chem.* 1996, 42 (10), 1589–1600.
- (67) Cox D. W.; Andrews B. J.; Wills D. E. Genetic Polymorphism of Alpha 2HS-Glycoprotein. *Am. J. Hum. Genet.* 1986, 38 (5), 699–706.
- (68) Tamara, S.; Franc, V.; Heck, A. J. R. A Wealth of Genotype-Specific Proteoforms Fine-Tunes Hemoglobin Scavenging by Haptoglobin. *Proc Natl Acad Sci. U S A* 2020, <https://doi.org/10.1073/pnas.2002483117>.
- (69) Thaysen-Andersen, M.; Packer, N. H.; Schulz, B. L. Maturing Glycoproteomics Technologies Provide Unique Structural Insights into the N-Glycoproteome and Its Regulation in Health and Disease. *Mol. Cell. Proteomics.* 2016, 15 (6), 1773–1790.
- (70) Li QK, Gabrielson E, Z. H. Application of Glycoproteomics for the Discovery of Biomarkers in Lung Cancer. *Proteomics Clin Appl.* 2012, 6 (5-6), 244–256.
- (71) Peracaula, R.; Barrabés, S.; Sarrats, A.; Rudd, P. M.; De Llorens, R. Altered Glycosylation in Tumours Focused to Cancer Diagnosis. *Dis. Markers*, 2008, 25 (4-5), 207–218.
- (72) Aoyagi, Y.; Suzuki, Y.; Igarashi, K.; Yokota, T.; Mori, S.; Suda, T.; Naitoh, A.; Isemura, M.; Asakura, H. Highly Enhanced Fucosylation of Alpha-Fetoprotein in Patients with Germ Cell Tumor. *Cancer.* 1993, 72 (2), 615–618.
- (73) Kölbl, A. C.; Andergassen, U.; Jeschke, U. The Role of Glycosylation in Breast Cancer Metastasis and Cancer Control. *Front. Oncol.* 2015, 5, 219.
- (74) Jrgensen, T.; Berner, A.; Kaalhus, O.; Tveter, K. J.; Danielsen, E.; Bryne, M. Up-Regulation of the Oligosaccharide Sialyl LewisX: A New Prognostic Parameter in Metastatic Prostate Cancer. *Cancer Res.* 1995, 55 (9), 1817–1819.
- (75) Stefaniuk, P.; Cianciara, J.; Wiercinska-Drapalo, A. Present and Future Possibilities for Early Diagnosis of Hepatocellular Carcinoma. *World J Gastroenterol.* 2010, 16 (4), 418–424.
- (76) Gornik, O.; Royle, L.; Harvey, D. J.; Radcliffe, C. M.; Saldova, R.; Dwek, R. A.; Rudd, P.; Lauc, G. Changes of Serum Glycans during Sepsis and Acute Pancreatitis. *Glycobiology.* 2007, 17 (12), 1321–1332.

- (77) Sillanauke, P.; Pönniö, M.; Jääskeläinen, I. P. Occurrence of Sialic Acids in Healthy Humans and Different Disorders. *Eur. J. Clin. Invest.* 1999, 29 (5), 413-425.
- (78) Sears, P.; Wong, C. H. Enzyme Action in Glycoprotein Synthesis. *Cell. Mol. Life Sci.* 1998, 54 (3), 223-252.
- (79) Piagnerelli, M.; Boudjeltia, K. Z.; Nuyens, V.; De Backer, D.; Su, F.; Wang, Z.; Vincent, J. L.; Vanhaeverbeek, M. Rapid Alterations in Transferrin Sialylation during Sepsis. *Shock.* 2005, 24 (1), 48-52.
- (80) Wang, F. L.; Cui, S. X.; Sun, L. P.; Qu, X. J.; Xie, Y. Y.; Zhou, L.; Mu, Y. L.; Tang, W.; Wang, Y. S. High Expression of α 2, 3-Linked Sialic Acid Residues Is Associated with the Metastatic Potential of Human Gastric Cancer. *Cancer Epidemiol.* 2009, 32 (5-6), 437-443.
- (81) Büll, C.; Stoel, M. A.; Den Brok, M. H.; Adema, G. J. Sialic Acids Sweeten a Tumor's Life. *Cancer Res.* 2014, 74 (12), 3199-3204.
- (82) Zhuo, Y.; Bellis, S. L. Emerging Role of alpha2,6-sialic Acid as a Negative Regulator of Galectin Binding and Function. *J. Biol. Chem.* 2011, 286 (8), 5935-5941.
- (83) Hsu, D. K.; Yang, R. Y.; Liu, F. T. Galectins in Apoptosis. *Methods Enzymol.* 2006, 417 (06), 256-273.
- (84) Schultz, M. J.; Swindall, A. F.; Bellis, S. L. Regulation of the Metastatic Cell Phenotype by Sialylated Glycans. *Cancer Metastasis Rev.* 2012, 31 (3-4), 501-518.
- (85) Lagana, A.; Goetz, J. G.; Cheung, P.; Raz, A.; Dennis, J. W.; Nabi, I. R. Galectin Binding to Mgat5-Modified N-Glycans Regulates Fibronectin Matrix Remodeling in Tumor Cells. *Mol. Cell. Biol.* 2006, 26 (8), 3181-3193.
- (86) Yamashita, T.; Forgues, M.; Wang, W.; Jin, W. K.; Ye, Q.; Jia, H.; Budhu, A.; Zanetti, K. A.; Chen, Y.; Qin, L. X.; et al. EpCAM and α -Fetoprotein Expression Defines Novel Prognostic Subtypes of Hepatocellular Carcinoma. *Cancer Res.* 2008, 68 (5), 1451-1461.
- (87) Lau, K. S.; Partridge, E. A.; Grigorian, A.; Silvescu, C. I.; Reinhold, V. N.; Demetriou, M.; Dennis, J. W. Complex N-Glycan Number and Degree of Branching Cooperate to Regulate Cell Proliferation and Differentiation. *Cell.* 2007, 129 (1), 123-134.
- (88) Shirato, K.; Nakajima, K.; Korekane, H.; Takamatsu, S.; Gao, C.; Angata, T.; Ohtsubo, K.; Taniguchi, N. Hypoxic Regulation of Glycosylation via the N-Acetylglucosamine Cycle. *J. Clin. Biochem. Nutr.* 2010, 48 (1), 20-25.
- (89) Granovsky, M.; Fata, J.; Pawling, J.; Muller, W. J.; Khokha, R.; Dennis, J. W. Suppression of Tumor Growth and Metastasis in Mgat5-Deficient Mice. *Nat. Med.* 2000, 6 (3), 306-312.
- (90) Tan, Z.; Wang, C.; Li, X.; Guan, F. Bisecting N-Acetylglucosamine Structures Inhibit Hypoxia-Induced Epithelial-Mesenchymal Transition in Breast Cancer Cells. *Front. Psychiatry.* 2018, 9, 210.
- (91) Dalziel, M.; Crispin, M.; Scanlan, C. N.; Zitzmann, N.; Dwek, R. A. Emerging Principles for the Therapeutic Exploitation of Glycosylation. *Science.* 2014, 343 (166), 1235681.
- (92) Hong, Q.; Ruhaak, L. R.; Stroble, C.; Parker, E.; Huang, J.; Maverakis, E.; Lebrilla, C. B. A Method for Comprehensive Glycosite-Mapping and Direct Quantitation of Serum Glycoproteins. *J. Proteome Res.* 2015, 14 (12), 5179-5192.

- (93) Riley, N. M.; Hebert, A. S.; Westphall, M. S.; Coon, J. J. Capturing Site-Specific Heterogeneity with Large-Scale N-Glycoproteome Analysis. *Nat. Commun.* 2019, 10 (1), 1311.
- (94) Suttapitugsakul, S.; Sun, F.; Wu, R. Recent Advances in Glycoproteomic Analysis by Mass Spectrometry. *Anal. Chem.* 2019, 92, 267–291.
- (95) Wu, D.; Li, J.; Struwe, W. B.; Robinson, C. V. Probing N-Glycoprotein Microheterogeneity by Lectin Affinity Purification-Mass Spectrometry Analysis. *Chem Sci.* 2019, 10 (19), 5146–5155.
- (96) Cao, L.; Diedrich, J. K.; Kulp, D. W.; Pauthner, M.; He, L.; Park, S. K. R.; Sok, D.; Su, C. Y.; Delahunty, C. M.; Menis, S.; et al. Global Site-Specific N-Glycosylation Analysis of HIV Envelope Glycoprotein. *Nat. Commun.* 2017, 8, 1–13.
- (97) Hu, M.; Lan, Y.; Lu, A.; Ma, X.; Zhang, L. Glycan-Based Biomarkers for Diagnosis of Cancers and Other Diseases: Past, Present, and Future. *Prog Mol Biol Transl Sci.* 2019, 162, 1–24.
- (98) Kailemia, M. J.; Park, D.; Lebrilla, C. B. Glycans and Glycoproteins as Specific Biomarkers for Cancer. *Anal. Bioanal. Chem.* 2017, 409 (2), 395–410.
- (99) Rudman, N.; Gornik, O.; Lauc, G. Altered N-Glycosylation Profiles as Potential Biomarkers and Drug Targets in Diabetes. *FEBS Lett.* 2019, 593 (13), 1598–1615.
- (100) Trbojević-Akmačić I, Vilaj M, Lauc G. High-Throughput Analysis of Immunoglobulin G Glycosylation. *Expert Rev. Proteomics.* 2016, 13 (5), 523–534.
- (101) Parker, B. L.; Thaysen-Andersen, M.; Fazakerley, D. J.; Holliday, M.; Packer, N. H.; James, D. E. Terminal Galactosylation and Sialylation Switching on Membrane Glycoproteins upon TNF-Alpha-Induced Insulin Resistance in Adipocytes. *Mol. Cell. Proteomics.* 2016, 15 (1), 141–153.
- (102) Kelleher, N. L. Top-down Proteomics. *Anal Chem.* 2004, 76 (11), 197A–203A.
- (103) Albanese, P.; Pagliano, C.; Tamara, S.; Saracco, G.; Scheltema, R. A. How Paired PSII-LHCII Supercomplexes Mediate the Stacking of Plant Thylakoid Membranes Unveiled by Structural Mass-Spectrometry. *Nat Commun.* 2020, 11 (1), 1361.
- (104) Siuti, N.; Kelleher, N. L. Decoding Protein Modifications Using Top-down Mass Spectrometry. *Nat Methods.* 2007, 4 (10), 817–821.
- (105) Čaval, T.; Tian, W.; Yang, Z.; Clausen, H.; Heck, A. J. R. Direct Quality Control of Glycoengineered Erythropoietin Variants. *Nat. Commun.* 2018, 9 (1), 3342.
- (106) Wu, D.; Struwe, W. B.; Harvey, D. J.; Ferguson, M. A. J.; Robinson, C. V. N-Glycan Microheterogeneity Regulates Interactions of Plasma Proteins. *Proc. Natl. Acad. Sci. U. S. A.* 2018, 115 (35), 8763–8768.
- (107) Yen, H.; Hopper, J. T. S.; Liko, I.; Allison, T. M.; Zhu, Y.; Wang, D.; Stegmann, M.; Mohammed, S.; Wu, B.; Robinson, C. V. Ligand Binding to a G Protein-Coupled Receptor Captured in a Mass Spectrometer. *Sci Adv.* 2017, 3 (6), e1701016.
- (108) Habberger, M.; Leiss, M.; Heidenreich, A.; Hafenmair, G.; Hook, M.; Bonnington, L.; Wegele, H.; Haindl, M.; Reusch, D.; Bulau, P.; et al. Rapid Characterization of Biotherapeutic Proteins by Size-Exclusion Chromatography Coupled to Native Mass Spectrometry. *MABs.* 2016, 8 (2), 331–339.

- (109) Ren, X.; Bai, H.; Pan, Y.; Tong, W.; Qin, P.; Yan, H. A Graphene Oxide-Based Immobilized PNGase F Reagent for Highly Efficient N-Glycan Release and MALDI-TOF MS Profiling. *Anal. Methods*. 2014, 6 (8), 2518.
- (110) Han, L.; Costello, C. E. Mass Spectrometry of Glycans. *Biochem.* 2013, 78 (7), 710-720.
- (111) Matsumoto, T.; Hatakeyama, S.; Yoneyama, T.; Tobisawa, Y.; Ishibashi, Y.; Yamamoto, H.; Yoneyama, T.; Hashimoto, Y.; Ito, H.; Nishimura, S. I.; et al. Serum N-Glycan Profiling Is a Potential Biomarker for Castration-Resistant Prostate Cancer. *Sci. Rep.* 2019, 9 (1), 1-8.
- (112) Reiding, K. R.; Bondt, A.; Hennig, R.; Gardner, R. A.; Flaherty, R. O. High-Throughput Serum N-Glycomics: Method Comparison and Application to Study Rheumatoid Arthritis and Pregnancy-Associated Changes. *Mol Cell Proteomics*. 2018, 18 (1), 3-15.
- (113) Hafkenscheid, L.; Bondt, A.; Scherer, H. U.; Huizinga, T. W. J.; Wuhler, M.; Toes, R. E. M.; Rombouts, Y. Structural Analysis of Variable Domain Glycosylation of Anti-Citrullinated Protein Antibodies in Rheumatoid Arthritis Reveals the Presence of Highly Sialylated Glycans. *Mol. Cell. Proteomics*. 2017, 16 (2), 278-287.
- (114) Tretter V.; Altmann F.; März L. Peptide-N4-(N-acetyl-beta-glucosaminyl)asparagine amidase F cannot release glycans with fucose attached alpha 1-3 to the asparagine-linked N-acetylglucosamine residue. *Eur. J. Biochem.* 1991, 199 (3), 647-652.
- (115) Liu, Z.; Cao, J.; He, Y.; Qiao, L.; Xu, C.; Lu, H.; Yang, P. Tandem 18O Stable Isotope Labeling for Quantification of N-Glycoproteome. *J. Proteome Res.* 2010, 9 (1), 227-236.
- (116) Cao, L.; Diedrich, J. K.; Ma, Y.; Wang, N.; Pauthner, M.; Park, S. K. R.; Delahunty, C. M.; McLellan, J. S.; Burton, D. R.; Yates, J. R.; et al. Global Site-Specific Analysis of Glycoprotein N-Glycan Processing. *Nat. Protoc.* 2018, 13 (6), 1196-1212.
- (117) Tep, S.; Hincapie, M.; Hancock, W. S. A MALDI-TOF MS Method for the Simultaneous and Quantitative Analysis of Neutral and Sialylated Glycans of CHO-Expressed Glycoproteins. *Carbohydr. Res.* 2012, 347 (1), 121-129.
- (118) Cech, N. B.; Enke, C. G. Effect of Affinity for Droplet Surfaces on the Fraction of Analyte Molecules Charged during Electrospray Droplet Fission. *Anal. Chem.* 2001, 73 (19), 4632-4639.
- (119) Marinas, M.; Sa, E.; Rojas, M. M.; Moalem, M.; Urbano, F. J.; Guillou, C.; Rallo, L. A nuclear magnetic resonance (1H and 13C) and isotope ratio mass spectrometry (delta13C, delta2H and delta18O) study of Andalusian olive oils. *Rapid Commun. Mass Spectrom.* 2010, 24, 1457-1466.
- (120) Lauber, M. A.; Yu, Y.; Brousmiche, D. W.; Hua, Z.; Koza, S. M.; Magnelli, P.; Guthrie, E.; Taron, C. H.; Fountain, K. J. Rapid Preparation of Released N-Glycans for HILIC Analysis Using a Labeling Reagent That Facilitates Sensitive Fluorescence and ESI-MS Detection. *Anal. Chem.* 2015, 87 (10), 5401-5409.
- (121) Wang, C.; Lu, Y.; Han, J.; Jin, W.; Li, L.; Zhang, Y.; Song, X.; Huang, L.; Wang, Z. Simultaneous Release and Labeling of O- and N-Glycans Allowing for Rapid Glycomic Analysis by Online LC-UV-ESI-MS/MS. *J. Proteome Res.* 2018, 17 (7), 2345-2357.
- (122) Lamari, F. N.; Kuhn, R.; Karamanos, N. K. Derivatization of Carbohydrates for Chromatographic, Electrophoretic and Mass Spectrometric Structure Analysis. *J Chromatogr B Anal. Technol Biomed Life Sci.* 2003, 793 (1), 15-36.

- (123) Reiding, K. R.; Blank, D.; Kuijper, D. M.; Deelder, A. M.; Wuhrer, M. High-Throughput Profiling of Protein N-Glycosylation by MALDI-TOF-MS Employing Linkage-Specific Sialic Acid Esterification. *Anal. Chem.* 2014, 86 (12), 5784–5793.
- (124) Sutton, C. W.; O'Neill, J. A.; Cottrell, J. S. Site-Specific Characterization of Glycoprotein Carbohydrates by Exoglycosidase Digestion and Laser Desorption Mass Spectrometry. *Anal. Biochem.* 1994, 218 (1), 34–46.
- (125) Nwosu C. C.; Strum J. S.; An H. J.; Lebrilla C. B. Enhanced Detection and Identification of Glycopeptides in Negative Ion Mode Mass Spectrometry. *Anal. Chem.* 2010, 82 (23), 9654–9663.
- (126) Medzilhradszky, K. F.; Kaasik, K.; Chalkley, R. J. Tissue-Specific Glycosylation at the Glycopeptide Level. *Mol. Cell. Proteomics.* 2015, 14 (8), 2103–2110.
- (127) Selman, M. H. J.; Hemayatkar, M.; Deelder, M.; Wuhrer, M. Cotton HILIC SPE Microtips for Microscale Purification and Enrichment of Glycans and Glycopeptides. *Anal. Chem.* 2011, 83 (7), 2492–2499.
- (128) Yan, J.; Li, X.; Yu, L.; Jin, Y.; Zhang, X.; Xue, X. Selective Enrichment of Glycopeptides/ Phosphopeptides Using Porous Titania Microspheres. *Chem. Commun.* 2010, 46 (30), 5488–5490.
- (129) Cao, W.; Huang, J.; Jiang, B.; Gao, X.; Yang, P. Highly Selective Enrichment of Glycopeptides Based on Zwitterionically Functionalized Soluble Nanopolymers. *Sci. Rep.* 2016, 6, 29776.
- (130) Sun, S.; Shah, P.; Eshghi, S. T.; Yang, W.; Trikanad, N.; Yang, S.; Chen, L.; Aiyetan, P.; Höti, N.; Zhang, Z.; et al. Comprehensive Analysis of Protein Glycosylation by Solid-Phase Extraction of N-Linked Glycans and Glycosite-Containing Peptides. *Nat. Biotechnol.* 2016, 34 (1), 84–88.
- (131) Woo, C. M.; Iavarone, A. T.; Spiciarich, D. R.; Palaniappan, K. K.; Bertozzi, C. R. Isotope-Targeted Glycoproteomics (IsoTaG): A Mass-Independent Platform for Intact N- and O-Glycopeptide Discovery and Analysis. *Nat. Methods.* 2015, 12 (6), 561–567.
- (132) Woo, C. M.; Felix, A.; Byrd, W. E.; Zuegel, D. K.; Ishihara, M.; Azadi, P.; Iavarone, A. T.; Pitteri, S. J.; Bertozzi, C. R. Development of IsoTaG, a Chemical Glycoproteomics Technique for Profiling Intact N- and O-Glycopeptides from Whole Cell Proteomes. *J. Proteome Res.* 2017, 16 (4), 1706–1718.
- (133) Qu, Y.; Sun, L.; Zhu, G.; Zhang, Z.; Peuchen, E. H.; Dovichi, N. J. Sensitive and Fast Characterization of Site-Specific Protein Glycosylation with Capillary Electrophoresis Coupled to Mass Spectrometry. *Talanta.* 2018, 179, 22–27.
- (134) Zhu, F.; Trinidad, J. C.; Clemmer, D. E. Glycopeptide Site Heterogeneity and Structural Diversity Determined by Combined Lectin Affinity Chromatography/IMS/CID/MS Techniques. *J. Am. Soc. Mass Spectrom.* 2015, 26 (7), 1092–1102.
- (135) Mookherjee, A.; Guttman, M. Bridging the Structural Gap of Glycoproteomics with Ion Mobility Spectrometry. *Curr. Opin. Chem. Biol.* 2018, 42, 86–92.
- (136) Pagel, K.; Harvey, D. J. Ion Mobility-Mass Spectrometry of Complex Carbohydrates: Collision Cross Sections of Sodiated N-Linked Glycans. *Anal. Chem.* 2013, 85 (10), 5138–5145.
- (137) Both, P.; Green, A. P.; Gray, C. J.; Šardžik, R.; Voglmeir, J.; Fontana, C.; Austeri, M.; Rejzek, M.; Richardson, D.; Field, R. A.; et al. Discrimination of Epimeric Glycans and

Glycopeptides Using IM-MS and Its Potential for Carbohydrate Sequencing. *Nat. Chem.* 2014, 6 (1), 65–74.

(138) Shi, T.; Fillmore, T. L.; Sun, X.; Zhao, R.; Schepmoes, A. A.; Hossain, M.; Xie, F.; Wu, S.; Kim, J. S.; Jones, N.; et al. Antibody-Free, Targeted Mass-Spectrometric Approach for Quantification of Proteins at Low Picogram per Milliliter Levels in Human Plasma/Serum. *Proc. Natl. Acad. Sci. U. S. A.* 2012, 109 (38), 15395–15400.

(139) Creese, A. J.; Cooper, H. J. Separation and Identification of Isomeric Glycopeptides by High Field Asymmetric Waveform Ion Mobility Spectrometry. *Anal. Chem.* 2012, 84 (5), 2597–2601.

(140) Chu, I. K.; Siu, C.; Lau, J. K.; Kit, W.; Mu, X.; Kuen, C.; Guo, X.; Wang, X.; Li, N.; Xia, Y.; et al. Proposed Nomenclature for Peptide Ion Fragmentation. *Int. J. Mass Spectrom.* 2015, 390, 24–27.

(141) Domon, B.; Costello, C. E. A Systematic Nomenclature for Carbohydrate Fragmentations in FAB-MS/MS Spectra of Glycoconjugates. *Glycoconj. J.* 1988, 5, 397–409.

(142) Marinas, M.; Sa, E.; Rojas, M. M.; Moalem, M.; Urbano, F. J.; Guillou, C.; Rallo, L. Characterizing Protein Glycosylation Sites through Higher-Energy C-Trap Dissociation. *Rapid Commun. Mass Spectrom.* 2010, 24, 1457–1466.

(143) Riley, N. M.; Coon, J. J. The Role of Electron Transfer Dissociation in Modern Proteomics. *Anal. Chem.* 2018, 90 (1), 40–64.

(144) Cao, L.; Tolić, N.; Qu, Y.; Meng, D.; Zhao, R.; Zhang, Q.; Moore, R. J.; Zink, E. M.; Lipton, M. S.; Paša-Tolić, L.; et al. Characterization of Intact N- and O-Linked Glycopeptides Using Higher Energy Collisional Dissociation. *Anal. Biochem.* 2014, 452 (1), 96–102.

(145) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *J. Am. Chem. Soc.* 1998, 120 (13), 3265–3266.

(146) Coon, J. J.; Shabanowitz, J.; Hunt, D. F.; Syka, J. E. P. Electron Transfer Dissociation of Peptide Anions. *J. Am. Soc. Mass Spectrom.* 2005, 16 (6), 880–882.

(147) Mirgorodskaya, E.; Roepstorff, P.; Zubarev, R. A. Localization of O-Glycosylation Sites in Peptides by Electron Capture Dissociation in a Fourier Transform Mass Spectrometer. *Anal. Chem.* 1999, 71 (20), 4431–4436.

(148) Adamson, J. T.; Håkansson, K. Infrared Multiphoton Dissociation and Electron Capture Dissociation of High-Mannose Type Glycopeptides. *J. Proteome Res.* 2006, 5 (3), 493–501.

(149) Volmer, Y. Q. and D. A. Electron-Based Fragmentation Methods in Mass Spectrometry: An Overview. *Mass Spectrom Rev.* 2017, 36 (1), 4–15.

(150) Frese, C. K.; Altelaar, A. F. M.; Van Den Toorn, H.; Nolting, D.; Griep-Raming, J.; Heck, A. J. R.; Mohammed, S. Toward Full Peptide Sequence Coverage by Dual Fragmentation Combining Electron-Transfer and Higher-Energy Collision Dissociation Tandem Mass Spectrometry. *Anal. Chem.* 2012, 84 (22), 9668–9673.

(151) Yu, Q.; Wang, B.; Chen, Z.; Urabe, G.; Glover, M. S.; Shi, X.; Guo, L. W.; Kent, K. C.; Li, L. Electron-Transfer/Higher-Energy Collision Dissociation (ET_hCD)-Enabled Intact Glycopeptide/Glycoproteome Characterization. *J. Am. Soc. Mass Spectrom.* 2017, 28 (9), 1751–1764.

- (152) Chen, Z.; Yu, Q.; Hao, L.; Liu, F.; Johnson, J.; Tian, Z.; Kao, W. J.; Xu, W.; Li, L. Site-Specific Characterization and Quantitation of N-Glycopeptides in PKM2 Knockout Breast Cancer Cells Using DiLeu Isobaric Tags Enabled by Electron-Transfer/Higher-Energy Collision Dissociation (EThcD). *Analyst*. 2018, 143 (11), 2508–2519.
- (153) Nilsson, J. Liquid Chromatography-Tandem Mass Spectrometry-Based Fragmentation Analysis of Glycopeptides. *Glycoconj. J.* 2016, 33 (3), 261–272.
- (154) Hinneburg, H.; Stavenhagen, K.; Schweiger-Hufnagel, U.; Pengelley, S.; Jabs, W.; Seeberger, P. H.; Silva, D. V.; Wührer, M.; Kolarich, D. The Art of Destruction: Optimizing Collision Energies in Quadrupole-Time of Flight (Q-TOF) Instruments for Glycopeptide-Based Glycoproteomics. *J. Am. Soc. Mass Spectrom.* 2016, 27 (3), 507–519.
- (155) Liu, M. Q.; Zeng, W. F.; Fang, P.; Cao, W. Q.; Liu, C.; Yan, G. Q.; Zhang, Y.; Peng, C.; Wu, J. Q.; Zhang, X. J.; et al. PGlyco 2.0 Enables Precision N-Glycoproteomics with Comprehensive Quality Control and One-Step Mass Spectrometry for Intact Glycopeptide Identification. *Nat. Commun.* 2017, 8 (1), 1–14.
- (156) Zeng, W. F.; Liu, M. Q.; Zhang, Y.; Wu, J. Q.; Fang, P.; Peng, C.; Nie, A.; Yan, G.; Cao, W.; Liu, C.; et al. PGlyco: A Pipeline for the Identification of Intact N-Glycopeptides by Using HCD- and CID-MS/MS and MS3. *Sci. Rep.* 2016, 6, 25102.
- (157) Reiding, K. R.; Bondt, A.; Franc, V.; Heck, A. J. R. The Benefits of Hybrid Fragmentation Methods for Glycoproteomics. *TrAC-Trends Anal. Chem.* 2018, 108, 260–268.
- (158) Bush, D. R.; Zang, L.; Belov, A. M.; Ivanov, A. R.; Karger, B. L. High Resolution CZE-MS Quantitative Characterization of Intact Biopharmaceutical Proteins: Proteoforms of Interferon- β 1. *Anal. Chem.* 2016, 88 (2), 1138–1146.
- (159) Haselberg, R.; Jong, G. J. De; Somsen, G. W. Low-Flow Sheathless Capillary Electrophoresis-Mass Spectrometry for Sensitive Glycoform Profiling of Intact Pharmaceutical Proteins. *Anal. Chem.* 2013, 85 (4), 2289–2296.
- (160) Nagel, T.; Meyer, B. Simultaneous Characterization of Sequence Polymorphisms, Glycosylation and Phosphorylation of Fibrinogen in a Direct Analysis by LC-MS. *Biochim. Biophys. Acta-Proteins Proteomics.* 2014, 1844 (12), 2284–2289.
- (161) Cleland, T. P.; Dehart, C. J.; Fellers, R. T.; Vannispén, A. J.; Greer, J. B.; LeDuc, R. D.; Parker, W. R.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S. High-Throughput Analysis of Intact Human Proteins Using UVPD and HCD on an Orbitrap Mass Spectrometer. *J. Proteome Res.* 2017, 16 (5), 2072–2079.
- (162) Waterbeemd, M. Van De; Tamara, S.; Fort, K. L.; Damoc, E.; Franc, V.; Bieri, P.; Itten, M.; Makarov, A.; Ban, N.; Heck, A. J. R. Dissecting Ribosomal Particles throughout the Kingdoms of Life Using Advanced Hybrid Mass Spectrometry Methods. *Nat. Commun.* 2018, 9 (1), 2493.
- (163) Rosati, S.; Yang, Y.; Barendregt, A.; Heck, A. J. R. Detailed Mass Analysis of Structural Heterogeneity in Monoclonal Antibodies Using Native Mass Spectrometry. *Nat. Protoc.* 2014, 9 (4), 967–976.
- (164) Van De Waterbeemd, M.; Fort, K. L.; Boll, D.; Reinhardt-Szyba, M.; Routh, A.; Makarov, A.; Heck, A. J. R. High-Fidelity Mass Analysis Unveils Heterogeneity in Intact Ribosomal Particles. *Nat. Methods.* 2017, 14 (3), 283–286.

- (165) Fort, K. L.; Waterbeemd, M. Van De; Boll, D.; Reinhardt-szyba, M.; Belov, M. E.; Sasaki, E.; Zschoche, R.; Hilvert, D.; Makarov, A. A.; Heck, A. J. R. Expanding the Structural Analysis Capabilities on an Orbitrap-Based Mass Spectrometer for Large Macromolecular Complexes. *Analyst*. 2018, 143 (1), 100-105.
- (166) Rosati, S.; Rose, R. J.; Thompson, N. J.; Van Duijn, E.; Damoc, E.; Denisov, E.; Makarov, A.; Heck, A. J. R. Exploring an Orbitrap Analyzer for the Characterization of Intact Antibodies by Native Mass Spectrometry. *Angew. Chemie-Int. Ed.* 2012, 51 (52), 12992-12996.
- (167) Yang, Y.; Wang, G.; Song, T.; Lebrilla, C. B.; Heck, A. J. R. Resolving the Micro-Heterogeneity and Structural Integrity of Monoclonal Antibodies by Hybrid Mass Spectrometric Approaches. *MAbs*. 2017, 9 (4), 638-645.
- (168) Dyachenko, A.; Wang, G.; Belov, M.; Makarov, A.; Jong, R. N. De; Bremer, E. T. J. Van Den; Parren, P. W. H. I.; Heck, A. J. R. Tandem Native Mass-Spectrometry on Antibody-Drug Conjugates and Submillion Da Antibody - Antigen Protein Assemblies on an Orbitrap EMR Equipped with a High-Mass Quadrupole Mass Selector. *Anal Chem*. 2015, 87 (12), 6095-6102.
- (169) Vimer S, Ben-Nissan G, S. M. Direct Characterization of Overproduced Proteins by Native Mass Spectrometry. *Nat. Protoc.* 2020, 15 (2), 236-265.
- (170) Bern, M.; Caval, T.; Kil, Y. J.; Tang, W.; Becker, C.; Carlson, E.; Kletter, D.; Sen, K. I.; Galy, N.; Hagemans, D.; et al. Parsimonious Charge Deconvolution for Native Mass Spectrometry. *J. Proteome Res.* 2018, 17 (3), 1216-1226.
- (171) Marty, M. T.; Baldwin, A. J.; Marklund, E. G.; Hochberg, G. K. A.; Benesch, J. L. P.; Robinson, C. V. Bayesian Deconvolution of Mass and Ion Mobility Spectra: From Binary Interactions to Polydisperse Ensembles. *Anal Chem*. 2015, 87 (8), 4370-4376.
- (172) Skala, W.; Wohlschlager, T.; Senn, S.; Huber, G. E.; Huber, C. G. MoFi-A Software Tool for Annotating Glycoprotein Mass Spectra by Integrating Hybrid Data from the Intact Protein and Glycopeptide Level. *Anal Chem*. 2018, 90 (9), 5728-5736.
- (173) Tong, W.; Wang, G. How Can Native Mass Spectrometry Contribute to Characterization of Biomacromolecular Higher-Order Structure and Interactions? *Methods*. 2018, 144, 3-13.
- (174) Yang, Y.; Liu, F.; Franc, V.; Halim, L. A.; Schellekens, H.; Heck, A. J. R. Hybrid Mass Spectrometry Approaches in Glycoprotein Analysis and Their Usage in Scoring Biosimilarity. *Nat. Commun.* 2016, 7, 13397.
- (175) Wohlschlager, T.; Scheffler, K.; Forstenlehner, I. C.; Skala, W.; Senn, S.; Damoc, E.; Holzmann, J.; Huber, C. G. Native Mass Spectrometry Combined with Enzymatic Dissection Unravels Glycoform Heterogeneity of Biopharmaceuticals. *Nat. Commun.* 2018, 9 (1), 1713.
- (176) Rose, R. J.; Damoc, E.; Denisov, E.; Makarov, A.; Heck, A. J. R. High-Sensitivity Orbitrap Mass Analysis of Intact Macromolecular Assemblies. *Nat. Methods*. 2012, 9 (11), 1084-1086.
- (177) Franc, V.; Zhu, J.; Heck, A. J. R. Comprehensive Proteoform Characterization of Plasma Complement Component C8 $\alpha\beta\gamma$ by Hybrid Mass Spectrometry Approaches. *J. Am. Soc. Mass Spectrom.* 2018, 29 (6), 1099-1110.

- (178) Franc, V.; Yang, Y.; Heck, A. J. R. Proteoform Profile Mapping of the Human Serum Complement Component C9 Revealing Unexpected New Features of N-, O-, and C-Glycosylation. *Anal. Chem.* 2017, 89 (6), 3483–3491.
- (179) Lin, Y. H.; Franc, V.; Heck, A. J. R. Similar Albeit Not the Same: In-Depth Analysis of Proteoforms of Human Serum, Bovine Serum, and Recombinant Human Fetuin. *J. Proteome Res.* 2018, 17 (8), 2861–2869.
- (180) Yang, Y.; Franc, V.; Heck, A. J. R. Glycoproteomics: A Balance between High-Throughput and In-Depth Analysis. *Trends Biotechnol.* 2017, 35 (7), 598–609.
- (181) Ikeda Y., Ihara H., Tsukamoto H., Gu J., Taniguchi N. Mannosyl (Beta-1,4)-Glycoprotein Beta-1,4-N-Acetylglucosaminyltransferase (MGAT3); β 1,4-N-Acetylglucosaminyltransferase III (GnT-III, GlcNAcT-III). *Handbook of Glycosyltransferases and Related Genes.* Springer, Tokyo. (https://doi.org/10.1007/978-4-431-54240-7_58)

Chapter 2

Similar albeit not the same; in-depth analysis of proteoforms of human serum, bovine serum and recombinant human fetuin

Yu-Hsien Lin^{1,2}, Vojtech Franc^{1,2} and Albert J.R. Heck^{1,2*}*

¹Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Padualaan 8, 3584 CH Utrecht, The Netherlands

²Netherlands Proteomics Center, Padualaan 8, 3584 CH Utrecht, The Netherlands

Authors for Correspondence: Vojtech Franc, v.franc@uu.nl (+31302536149) or Albert Heck, a.j.r.heck@uu.nl (+31302536797)

Based on:

Lin, Y. H.; Franc, V.; Heck, A. J. R. Similar Albeit Not the Same: In-Depth Analysis of Proteoforms of Human Serum, Bovine Serum, and Recombinant Human Fetuin. *J. Proteome Res.* 2018, 17 (8), 2861–2869

Abstract

Fetuin, also known as Alpha-2-Heremans Schmid Glycoprotein (AHSG), belongs to some of the most abundant glycoproteins secreted into the bloodstream. In blood, fetuins exhibit functions as carriers of metals and small molecules. Bovine fetuin, which harbors 3 N-glycosylation sites and a suggested half a dozen O-glycosylation sites, has been often used as a model glycoprotein to test novel analytical workflows in glycoproteomics. Here we characterize and compare fetuin in depth, using protein from three different biological sources; human serum, bovine serum and recombinant human fetuin expressed in HEK-293 cells, with the aim to elucidate similarities and differences between these proteins and the post-translational modifications they harbor. Combining data from high-resolution native mass spectrometry and glycopeptide centric LC-MS analysis, we qualitatively and quantitatively gather information on fetuin protein maturation, N-glycosylation, O-glycosylation, and phosphorylation. We provide direct experimental evidence that both the human serum and part of recombinant proteins are processed into two chains (A and B) connected by a single inter-chain disulfide bridge, whereas bovine fetuin remains a single chain protein. Although two N-glycosylation sites, one O-glycosylation site, and a phosphorylation site are conserved from bovine to human, the stoichiometry of the modifications and the specific glycoforms they harbor are quite distinct. Comparing serum and recombinant human fetuin, we observe that the serum protein harbors a much simpler proteoform profile, indicating that the recombinant protein is not ideally engineered to mimic human serum fetuin. Comparing the proteoform profile and post-translational modifications of human and bovine serum fetuin, we observe that although the gene-structure of these two proteins are alike, they represent quite distinct proteins when their glycoproteoform profile is also taken into consideration.

Keywords: fetuin / alpha-2-HS glycoprotein / glycoprotein / serum proteins / native mass spectrometry / glycopeptides / proteoforms / hybrid mass spectrometry / N-glycosylation / O-glycosylation

Introduction

The fetuins are a group of related proteins belonging to the cystatin superfamily. These multifunctional proteins were decades ago already identified in various mammals including humans^{1,2}. Fetuin was discovered in 1944 by Kai Pedersen in fetal calf serum³. There has been some initial confusion related to the fetuin naming, which led to the mixed use of the name alpha-2-HS glycoprotein in some species and fetuin in others. Since 1990 bovine fetuin and human alpha-2-HS glycoprotein have been considered as species homologs⁴. Here, for the naming we follow the recommendation by Brown et al. 1992 and use the name “fetuin” for both human fetuin (hFet) and bovine fetuin (bFet)⁵. Fetuins are known for their complicated heterogeneous structure and many reported discrepancies related to their putative biological functions. Despite years of research, the biological function of fetuins and a true understanding of their biological importance is still unclear. Some consensus has been found in the role of hFet in calcium metabolism^{6,7} and insulin signaling⁸. hFet is also extensively studied for its potential relevance as a metabolic biomarker^{9,10}. Increased levels of hFet have been linked to higher risk of cardiovascular disease (CVD) and incident type 2 diabetes (T2DM)^{11,12}. However, there are several obstacles preventing the use of hFet as a biomarker for those and any other diseases. These include for example, a lack of reference values, inconsistent values from various commercial enzyme-linked immunosorbent assays (ELISA) and the unknown effect of PTMs on hFet clinical measurements¹³. One source of confusion may also originate from some *in vitro* studies, where recombinant human fetuin was used^{14,15}. Serum hFet is predominantly synthesized in the liver, where its N-glycosylation pattern originates¹⁶. The glycosylation machinery is species specific and thus proteins produced by different expression systems provide products with distinct glycosylation patterns¹⁷⁻¹⁹. Recombinant human fetuin (rhFet) synthesized in human embryonic kidney cells (HEK-293) is often applied for antibody validation, ELISA assays, immunoprecipitation or protein functional assays. Therefore, we included this rhFet into our study to investigate potential structural differences between serum derived hFet and recombinant rhFet.

All described fetuins are glycoproteins and especially the glycosylation profile of bFet is well documented in the literature^{20,21}. For that reason, bFet has also been widely used as a standard glycoprotein for method development in glycoproteomics. Posttranslational modifications (PTMs) on hFet have been less described and even the primary structure of mature hFet is somehow elusive. Amino acid sequence alignment of hFet and bFet shows a relatively high sequence similarity (~70%) suggesting a high degree of similarity (Supplementary Figure 1). Also, fetuins from other mammalian species reflect a high sequence conservation showing 60-70% homology at the amino acid level and 80-90% homology at the cDNA level⁵. Nevertheless, mature hFet harbors some unique features, as it is present in serum in the form of two chains

connected to each other by a single inter-chain disulfide bridge while other fetuins, including bFet, are found in serum in a single-chain form²². PTMs of proteins, in general, play an important role in regulating their structure, function, and interactions^{23,24}. Regarding hFet and bFet, the published data document the presence of N- and O-glycosylation sites and a few phosphosites^{25,26}. Although the type of modifications is alike in both proteins, the number, structures, and distribution on their primary structure is distinct. The O-glycosylation sites are less conserved than the N-glycosylation sites⁴. Additionally, the number of reported phosphosites on bFet is higher in comparison to hFet^{20,25}. Differences in the structure of N-linked glycans released from fetuins isolated from various biological sources are well documented⁵. These findings supported the concept of species specificity of N-glycan structure in glycoproteins from different species^{27,28}. In this work, we follow up these earlier studies and extend them by an in-depth site-specific characterization of fetuins from three different biological sources using state-of-art hybrid mass spectrometry (MS) approaches. In our earlier works, we showed a great utility of combining high-resolution MS and peptide-centric MS for comprehensive and unbiased analysis of blood serum protein PTMs^{29,30,31}. Here we aim to provide detailed information illustrating the differences between hFet, bFet and rhFet, with an emphasis on their primary structure and PTMs. Our data provide new evidence for posttranslational events occurring on the three fetuins and show how alike gene products synthesized in various species can mature into very different molecules with potentially different functions.

Material and methods

Chemicals and materials.

hFet (Alpha-2-HS glycoprotein; Uniprot Code: P02765), bFet (Bovine fetuin; Uniprot Code: P12763) and rhFet (Recombinant Alpha-2-HS glycoprotein expressed in HEK293 cells), dithiothreitol (DTT), iodoacetamide (IAA), trifluoroacetic acid (TFA), ammonium bicarbonate (ABC) and ammonium acetate (AMAC) were purchased from Sigma-Aldrich (Steinheim, Germany). Acetonitrile was purchased from Biosolve (Valkenswaard, The Netherlands). Sequencing grade trypsin was obtained from Promega (Madison, WI). Gluc-C, Lys-C, PNGaseF³², and Sialidase were obtained from Roche (Indianapolis, USA). Alkaline phosphatase was purchased from New England Biolabs (Ipswich, MA).

Sample preparation for native MS

Unprocessed hFet, bFet and rhFet in deionized water, containing 25-30 μ g of the protein, were buffer exchanged into 150 mM aqueous ammonium acetate (AMAC) (pH 7.2) by ultrafiltration

(vivaspin500, Sartorius Stedim Biotech, Germany) with a 10 kDa cut-off filter. The protein concentration was measured by UV absorbance at 280 nm and adjusted to 2-3 μM prior to native MS analysis. The enzyme PNGase was used to remove the N-glycans of the fetuins and sialidase was used to cleave sialic acid residues³². Alkaline phosphatase was used for the removal of phosphate groups. DTT (4 mM) was used to reduce the disulfide bonds between the A chain and B chain in hFet. All samples for different treatments were buffer exchanged to 150 mM AMAC (pH 7.2) prior to native MS analysis.

Native MS analysis of hFet, bFet and rhFet

Samples were analyzed on a modified Exactive Plus Orbitrap instrument with extended mass range (EMR) (Thermo Fisher Scientific, Bremen) using a standard m/z range of 500-10,000, as described in detail previously³³. The voltage offsets on the transport multi-poles and ion lenses were manually tuned to achieve optimal transmission of protein ions at elevated m/z . Nitrogen was used in the HCD cell at a gas pressure of $6-8 \times 10^{-10}$ bar. The MS parameters were used typically: spray voltage 1.2-1.3 V, source fragmentation 30 V, source temperature 250 °C, collision energy 30 V, and resolution (at m/z 200) 17,500. The mass spectrometer was calibrated using CsI clusters as described previously³³.

Native MS data analysis

The accurate masses of observed hFet, bFet, and rhFet proteoforms were extracted by deconvoluting the ESI spectrum to zero-charge spectrum using Intact Mass software by Protein Metrics in ver. 1.5³⁴. For PTM composition analysis, data was processed manually and glycan structures were deduced on the basis of known biosynthetic pathways. The average masses were used for these calculations, including hexose/mannose/galactose (Hex/Man/Gal, 162.1424 Da), N-acetylhexosamine/N-acetylglucosamine (HexNAc/GlcNAc, 203.1950 Da), deoxyhexose (dHex, 146.1430 Da), N-acetylneuraminic acid (Neu5Ac, 291.2579 Da and phosphorylation (Pho 79.9799 Da). All used symbols and text nomenclature are based on the recommendation of the Consortium for Functional Glycomics³⁵.

In-solution digestion for peptide-centric proteomics

All proteins (bFet, hFet, and rhFet) were reconstituted in 50 mM ABC at a concentration of 1 mg/mL, then reduced with 4 mM DTT at 56 °C for 30 min and alkylated with 8 mM IAA at room temperature for 30 min in the dark. bFet was digested for 3 h with Glu-C at an enzyme-to-protein-ratio of 1:75 (w/w) at 37 °C and the resulted peptide mixtures were further treated

by using trypsin (1:100; w/w). hFet and rhFet were digested for 3 h with Lys-C at an enzyme-to-protein-ratio of 1:75 (w/w) at 37 °C and the resulted peptide mixtures were further treated by using Glu-C (1:100; w/w). All proteolytic digests containing modified glycopeptides were desalted by using GELoader tips filled with POROS Oligo R3 50 µm particles, dried, and reconstituted in 20 µL of 0.1% FA prior to liquid chromatography (LC)-MS and MS/MS analysis³⁶.

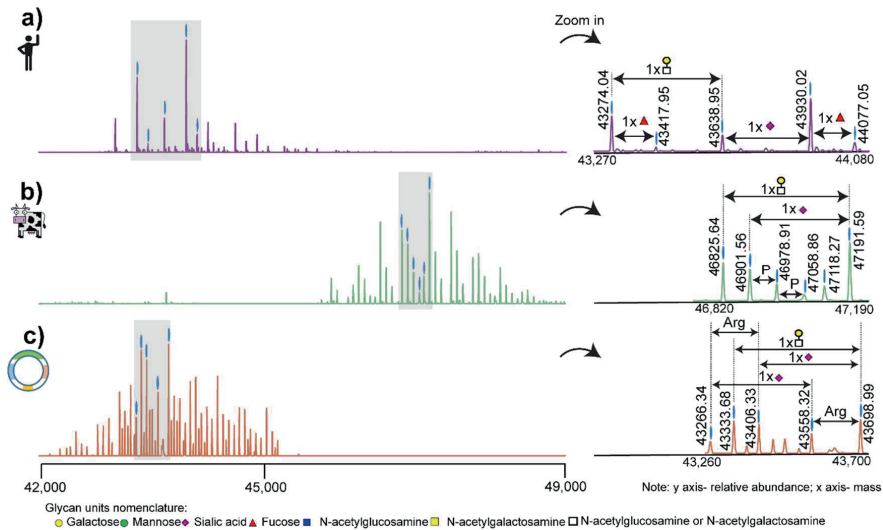


Figure 1. Deconvoluted zero-charge mass spectra of (a) hFet, (b) bFet and (c) rhFet. The zoom-ins on the right depict the most abundant peaks in the spectra to more clearly show the observed mass differences in each spectrum that originated mainly from distinct glycan moieties. The presence of a third N-glycosylation site in bFet increases the molecular weight and glycan heterogeneity of bFet compared to that of hFet. In (c), all proteoforms of rhFet are present in pairs, due to the co-occurrence of proteoforms with and without Arginine, making the spectrum twice as complex as that of hFet. The glycan nomenclature used is indicated at the bottom.

LC-MS and MS/MS analysis

All peptides generated from fetuin (typically 300 fmol) were separated and analyzed using an Agilent 1290 Infinity HPLC system (Agilent Technologies, Waldbronn, Germany) coupled on-line to an Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). Reversed-phase separation was accomplished using a 100 µm inner diameter 2 cm trap column (in-house packed with ReproSil-Pur C18-AQ, 3 µm) (Dr. Maisch GmbH, Ammerbuch-Entringen, Germany) coupled to a 50 µm inner diameter 50 cm analytical column (in-house packed with Poroshell 120 EC-C18, 2.7 µm) (Agilent Technologies, Amstelveen, The Netherlands). Mobile-phase solvent A consisted of 0.1% formic acid in water, and mobile-

phase solvent B consisted of 0.1% formic acid in acetonitrile. The flow rate was set to 300 nL/min. A 45 min gradient was used as follows: 0-5 min, 100% solvent A; 13-44% solvent B within 20 min; 44-100% solvent B within 3 min; 100% solvent B for 1 min; 100% solvent A for 17 min. For the MS scan, the mass range was set from 375-1500 m/z at a resolution of 120,000 and the AGC target was to 4×10^5 . For the MS/MS measurements, both higher-energy collision dissociation (HCD), and electron-transfer combined with higher-energy collision dissociation (ETHcD) were used and performed with normalized collision energy of 35%. For the MS/MS scan, the mass range was set from 125-2000 m/z ; the AGC target was set to 5×10^4 . The precursor isolation width was 1.6 Da, and the maximum injection time was set to 200 ms.

LC-MS and MS/MS data analysis

The raw data files were processed using Proteome Discoverer 2.2 software (Thermo Fisher Scientific) (PD 2.2) equipped with the Byonic software node (Protein Metrics Inc.)³⁷. The following parameters were used for data searches in Byonic: precursor ion mass tolerance, 10 ppm; product ion mass tolerance, 20 ppm; fixed modification, Cys carbamidomethyl; variable modification: Met oxidation, STY phosphorylation, and both N- and O- glycosylation from mammalian glycan databases; the allowed number of peptide miss-cleavages was set to 3. The protein database used contained the hFet (Uniprot Code: P02765) or bFet (Uniprot Code: P12763) amino acid sequences. Site-specific quantification of the fetuins PTMs was performed as follows. Each peptide that contains PTM sites was normalized individually so that the sum of all its proteoform areas was set to 100%. The average peptide ratios from all measurements were taken as a final estimation of the abundance. The XICs were obtained using the software Thermo Proteome Discoverer 2.2.0.388. The glycan structures of each glycoform were manually annotated. Hereby reported glycan structures are depicted without the linkage type of the glycan units since the acquired MS/MS patterns do not provide such information.

Combining native MS and peptide-centric proteomic data

Validation of the obtained proteoform profiles of all three fetuins was assessed by an integrative approach combining the native MS data with the peptide-centric proteomics data. This approach has been described in detail previously²⁹. Briefly, *in silico* data construction of the “intact protein spectra” was performed based on the masses and relative abundances of all site-specific PTMs derived from the peptide-centric analysis. Subsequently, the constructed spectra were compared to the experimental native MS spectra of the fetuins. The similarity between the two independent data sets (Native MS spectra and constructed spectra based on peptide-

centric data) was expressed by a Pearson correlation factor. All R scripts used for the spectra simulation are available at github (<https://github.com/Yang0014/glycoNativeMS>).

Results

Native MS reveals remarkable differences in structural heterogeneity among hFet, bFet and rhFet

We started our investigation by acquiring high-resolution native ESI-MS spectra of hFet, bFet and rhFet. Even at first glance, deconvoluted zero-charge spectra show remarkable differences among these three samples (Fig. 1). When recording the full proteoform profile of intact hFet by native MS, more than 30 peaks could be baseline resolved. This number is in pronounced contrast to the number of detected peaks in the native MS spectra of bFet (>40 peaks) and rhFet (>50 peaks). According to literature, the molecular heterogeneity of fetuins is mainly caused by N-glycosylation and O-glycosylation. Indeed, the observed mass differences among the most abundant peaks in all three native spectra of the fetuins correspond to the presence of glycans. Nevertheless, a closer look at the proteoform profiles reveals some other less expected structural variants. The heterogeneity of bFet native spectrum is significantly enriched by the presence of many lower intense peaks indicating the attachment of phosphate moieties (+80 Da) (Fig. 1b). The most complicated proteoform profile among all three samples can be observed for rhFet (Fig. 1c). Interestingly, many proteoform signals in rhFet co-exist in pairs differing from each other by a mass of 156 Da. This is likely due to the mass increment of Arginine, which will be discussed later.

Due to the high complexity of the data, we decided to focus on the full annotation of the native MS spectrum of hFet, and here refer to bFet or rhFet only in specific cases, also as the glycoproteome profile of bFet has already been well characterized^{20,38-40}. The protein backbone amino acid sequence of hFet represents an average mass of 37,177.01 Da. This mass was calculated based on the gene sequence of hFet lacking the N-terminal signal peptide, including the mass shifts induced by the 6 disulfide bonds and the absence of Arginine at position 322²². Determining the exact backbone mass allowed us to calculate a mass shift of 6751.81 Da induced by the PTMs on the most abundant peak in the hFet native MS spectrum (43,930.02 Da). Next, we enzymatically treated hFet attempting to remove either all N-glycans, the sialic acid moieties or phosphates, which results in specific mass shifts that we subsequently recorded by native MS. For the specific cleavage of N-glycosylations, we used PNGase F, sialidase for the removal of sialic acids and finally alkaline phosphatase for the release of phosphate residues. Incubation with PNGase F resulted in the removal of only one N-glycan (Supplementary Figure 2a). The mass difference of 2204 Da between the most abundant intact

hFet proteoform with 43,930.02 Da ($m/z = 3380.24$) and the N-deglycosylated hFet with 41,724.80 Da ($m/z = 3210.60$), indicated the attachment of a N-glycan with the carbohydrate composition of HexNAc₄Hex₅Neu5Ac₂. It is well known that hFet contains two N-glycosylation sites. However, even prolonged incubations with PNGase F did not lead to the complete removal of N-glycans under native conditions. This is a well-documented problem attributed to the lower accessibility of the second N-glycosylation site due to steric hindrance. Sialidase treatment of hFet resulted in a pronounced simplification of the structural heterogeneity of the hFet proteoforms (Supplementary Figure 2b), implying that the heterogeneity of hFet is mainly due to extensive modification with variable amounts of sialic acids. In total, 8 sialic acids were removed from the most abundant hFet proteoform as indicated by a mass shift of 2330 Da (8 x 291 Da). Lastly, we subjected hFet to treatment with alkaline phosphatase which resulted in the cleavage of one phosphate group from all hFet proteoforms (Supplementary Figure 2c). Although the composition of the second N-glycan on the most abundant hFet proteoform could not be determined due to the incomplete removal of N-glycans, the presence of this N-glycan is undoubtable based on the calculated PTM mass and information in literature¹⁶. The mass differences 365 (HexNAc₁Hex₁) and 656 Da (HexNAc₁Hex₁Neu5Ac₁) between the particular proteoforms correspond either to variability in the number of antennas on the N-glycans and/or the presence of O-glycans. Combining all this information, we can assume that the overall PTM composition of the most abundant hFet proteoform includes two N-glycans, several O-glycans, and one phosphate moiety.

Native MS of hFet treated with DTT reveals its two polypeptide chain structure

In addition to the structural variability originating from various PTMs on fetuins, the primary polypeptide architecture is another prominent origin of differences between hFet and bFet. Almost three decades ago, Kellermann et al. isolated hFet from fresh human serum in the presence of proteinase inhibitors, and determined that the major circulating form of hFet is likely a two polypeptide chain protein with a heavy chain (A-chain) of 321 residues and a light chain (B-chain) of 27 residues²² (Supplementary Figure 3). This circulating form of hFet contains a propeptide (also called connecting peptide) with a missing C-terminal Arginine residue (position 322) attached to the A-chain. The A-chain and B-chain are connected to each other by a single inter-chain disulfide bridge. We treated hFet with DTT to disrupt this linkage and validate the hypothesized arrangement of the primary structure. The subsequent recorded mass spectra, shown in Figure 2 reveal that the B-chain was released under reducing conditions from the A-chain and confirms the two polypeptide chain form of hFet. Notably, the released B-chain appeared to be, at least, in two structural variants, unmodified and modified with a

glycan (HexNAc₁Hex₁Neu5Ac₁). This confirms not only the existence of hFet in its two polypeptide form, but also the presence of one O-glycan on the B-chain.

Site-specific characterization of PTMs on hFet, bFet and rhFet by peptide-centric proteomics

Since the fetuins harbor at least three different types of PTMs, their analysis at the peptide level is a challenging task. We used two different combinations of proteolytic enzymes for the fetuin digestion. After a careful inspection of the fetuin amino acid sequences, we digested hFet and rhFet with Lys-C and Glu-C and bFet with trypsin and Glu-C. Combining tryptic/Lys-C with Glu-C specificity for the digestion of fetuins resulted in a set of peptides with a suitable length for subsequent sequencing by LC-MS/MS analysis. Lys-C was selected instead of trypsin for the digestion of hFet and rhFet to enable the confirmation of the absence/presence of the C-terminal Arginine on the A-chain. After enzymatic digestion, the peptide mixtures were subjected to ETHcD fragmentation to obtain extensive fragment ions

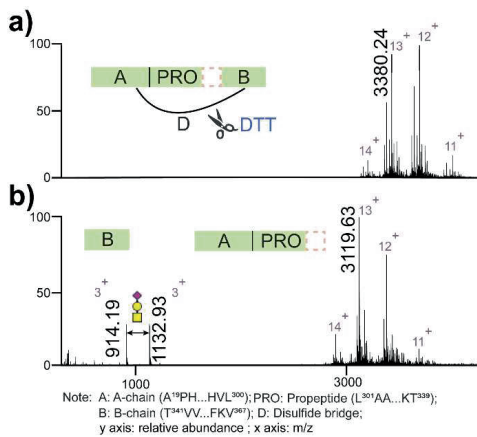


Figure 2. Full native ESI-MS spectra of intact hFet sprayed from aqueous ammonium acetate. **(a)** The schematic cartoon shows that the B-chain is connected to the A-chain by a disulfide bridge. **(b)** Full native ESI-MS spectra of hFet upon treatment by DTT. The released B-chain and A-chain are observed. The peaks at m/z of 914.19 and 1132.93 correspond to the B-chain and B-chain with 1 O-glycan (HexNAc₁Hex₁Neu5Ac₁), respectively. Comparing the most abundant peak on the charge state 13⁺ with m/z of 3380.24 in (a) and 3119.63 in (b), the mass difference indeed originates from the released B-chain harboring the O-glycan. The dashed line box represents missing C-terminal A-chain Arginine.

of both the glycan and the peptide moieties of glycopeptides. In addition to the PTM identification, we also assessed the relative abundances of the different (glyco/phospho)peptide isoforms. As a result, we identified and relatively quantified peptide isoforms from the putative N- and O-glycosylation and phosphorylation sites on all three investigated fetuins. The list of all modified peptide isoforms on hFet, bFet and rhFet and their relative quantification, based on XICs, can be found in Supplementary Table 1. Annotated HCD/ETHcD spectra of all glycopeptides observed for hFet are provided in the Supplementary data.

Comparison of the glycosylation profile of hFet, bFet and rhFet

A summary of the site-specific glycosylation patterns in all investigated fetuins are depicted in Figure 3. Focusing first on the N-glycosylation in hFet and bFet (Fig. 3a, b), we note that the three N-glycosylation consensus sequences in bFet are well-conserved in fetuins of most mammals. One exception is however hFet, which has the site N99 in bFet replaced by an Arginine, preventing its N-glycosylation in hFet. If we next compare the N-glycans on the other two conserved sites (N156 and N176), both bFet and hFet contain complex N-glycans, but differ in their structural composition and level of microheterogeneity. Sialylated diantennary complex type structures dominate on hFet and are typical for human serum proteins synthesized in the liver. Some less abundant glycoforms were found to be core fucosylated, which is in sharp contrast to bFet, where no fucosylation of N-glycans was observed at all. The N-glycans present on bFet show a higher degree of branching and also their quantitative distribution is more equal compared to the relatively more homogeneous hFet. In addition to this, approximately one-third of bFet molecules bear no glycan structure on the N176 site. Comparing next the N-glycosylation patterns on hFet and rhFet (Fig. 3a, c), we observe remarkable differences at both N156 and N176 N-glycosylation sites. The most noticeable difference is the observed extensive microheterogeneity on both N-glycosylation sites in rhFet, represented by a repertoire of complex core fucosylated glycan structures with a large variety of branches.

The heterogeneity of the fetuin glycosylation patterns is further increased by the presence of O-glycosylation. The O-glycopeptides identified and characterized in the present study cover all known hFet O-glycosylation sites and were used to determine the composition and occupancy of the attached O-glycans (Fig. 3a). As mentioned earlier, O-glycosylation sites are less conserved among fetuins, which also partly explains the observed major differences in O-glycosylation patterns. hFet/rhFet contain in total three reported O-glycosylation sites (T256, T270, and S346) while bFet harbors five sites (S271, T280, S282, S296, and S341). In all three investigated fetuins, we observed O-glycopeptides bearing simple mucin-type core 1 O-glycans

with one or two sialic acids. Regarding the two conserved O-glycosylation sites on hFet and bFet (Fig. 3a, b), T270 on hFet harbors mostly disialylated O-glycans while monosialylated structures reside on the analogous site S271 on bFet. The site S271 occurs in a cluster together with additional O-glycosylation sites T280, S282 and S296, which are absent in hFet. In both fetuins the second conserved O-glycosylation site is only partially occupied by glycosylation. S346 on hFet is unmodified in approximately 40% of the molecules and bears mostly monosialylated O-glycans. Occupancy of S341 on bFet is negligible²⁰. The last hFet O-glycosylation site T256 is not present on bFet and is almost fully occupied by monosialylated O-glycans. Differences in the O-glycosylation patterns between hFet and rhFet are rather marginal (Fig. 3a, c). The site T256 differs somewhat in the degree of sialylation, T270 is almost identical, and S346 on rhFet has seemingly a lower occupancy when compared to hFet.

Differences in phosphorylation of hFet, bFet and rhFet

The third type of PTM occurring on fetuins is phosphorylation. hFet contains two documented phosphorylation sites (S138 and S330) and bFet supposedly four (S138, S320, S323, and S325). Similar to the O-glycosylation sites, the phosphosites are less conserved among fetuins. hFet and bFet have two consensus phosphosites (S138 and S330/S325), however, their occupancy varies. While S138 on hFet was always found to be fully occupied, the analogous site on bFet was found to be occupied in only 10% of the proteoforms. The second phosphosite is located in a much less conserved region. In hFet, this sequence domain corresponds to the C-terminal A-chain propeptide and accommodates the partially phosphorylated site S330. The site S325 on bFet is situated in close proximity to the other two phosphosites, S320 and S323, and their occupancy is also only partial. Due to the low abundance of the phosphorylated peptides and phosphate lability, we were not able to unambiguously localize the neighboring phosphosites on bFet. Finally, in sharp contrast to hFet we did not find any evidence of phosphorylation on rhFet.

Data integration and major structural differences among hFet, bFet and rhFet

Having both the native MS data and peptide-centric data on all three fetuins available, we cross-validated the data to obtain a comprehensive view of the proteoform profiles of the fetuins. Figure 4 highlights the major differences observed among the most abundant proteoforms of hFet, bFet, and rhFet. In addition to the described structural differences based on various PTMs, hFet differs from bFet by its unique two polypeptide chain structure. The native MS data on rhFet suggested incomplete cleavage of the C-terminal Arginine at position 322. This

observation was further supported by two identified Lys-C peptides with the amino acid sequence $^{321}(\text{K})\text{TRTVVQPSVGAAGPVPVPPCPGRIRHFK}(\text{V})^{348}$ and $^{323}(\text{R})\text{TVVQPSVGAAGPVPVPPCPGRIRHFK}(\text{V})^{348}$, respectively (Supplementary Figure 4). From this data, we conclude that rhFet occurs as a mixture of a one- and two-chain polypeptide form, differing from each other by the presence/absence of the C-terminal Arginine at the A-chain.

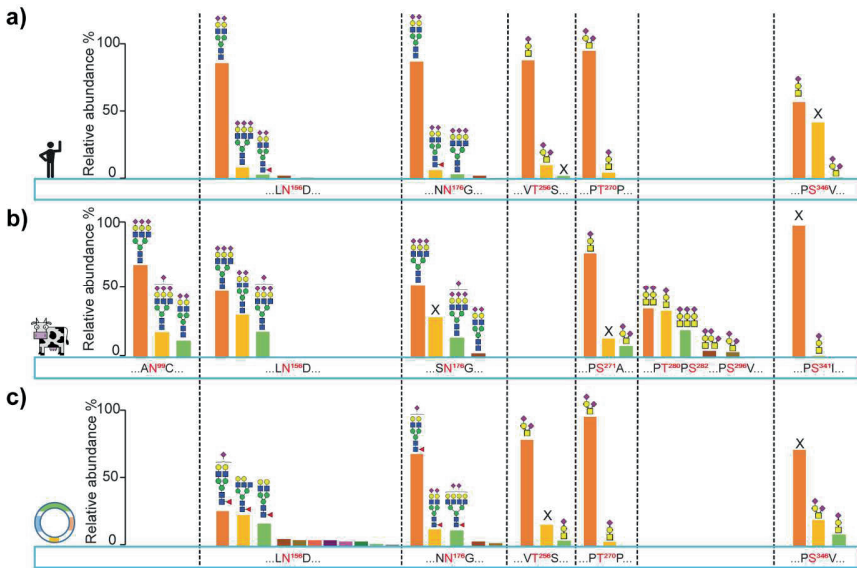


Figure 3. Overview of the observed qualitative and semi-quantitative site-specific glycosylation in (a) hFet, (b) bFet and (c) rhFet. The conserved glycosylation sites are depicted in the same column. Relative abundances of peptide proteoforms were estimated from their corresponding ion chromatograms (XICs). On a given modification site, the abundance of peptide proteoforms were normalized to 100%. All O-glycan structures and N-glycan structures attached to the 3 most abundant peptide isoforms of each site are depicted; further details of occupancy on each site is provided in the Supplementary Table 1. (X means unmodified).

Next, we cross-validated the peptide-centric and native MS data on hFet and bFet by a comparative comparison between the native MS spectra and an *in silico* constructed MS spectrum based on all the quantitative information we gathered from the LC-MS/MS peptide-centric data. For hFet and bFet, we achieved a high degree of correlation (~ 0.9) between our native MS and peptide-centric MS approach (Supplementary Figure 5). Therefore, all hFet and bFet species predicted from the peptide-centric data were filtered by taking 1 % cut-off in relative intensity of the peaks in the experimental native spectrum and mass deviations were manually checked. Applying these criteria resulted in a list containing 21 hFet and 33 bFet distinct proteoforms

(Supplementary Table 2). As an example of our data, we provide the fully annotated native MS spectrum of hFet demonstrating the (near) completeness of our analysis (Figure 5). Although we could explain most of the ion signals detected in the native MS spectra from hFet and bFet, some unmatched low abundant ion signals are still present. Those mostly correspond to adducts bearing Na⁺ and/or K⁺ ions, which represent frequent artifacts formed during the ESI ionization process.

Discussion

To disclose that similar genes can lead to a plethora of distinct and different proteoforms, we here meticulously analyzed and compared fetuin originating from three different biological sources, using an integrative MS approach allowing an all-inclusive analysis of protein PTMs. Both hFet and bFet PTMs have been the subject of several structural studies^{20,38-43}. However, there has been, as far as we know, no study describing data on all three types of modifications (*i.e.* N-glycosylation, O-glycosylation and phosphorylation) on hFet and bFet, pinpointing the major structural differences in a qualitative and quantitative site-specific manner. Earlier studies also reported on differences in fetuin DNA sequences, amino acid sequences, and PTMs amongst a range of species^{44,5,39,45}. Those studies provided first evidence for specific structural variabilities among fetuins. Most mammalian fetuins show a high degree of sequence conservation, but their final protein structure can be significantly altered by species specific PTMs. This intriguing phenomenon complicates structural and functional studies of proteins in general. Our main aim here was to demonstrate how three fetuins (hFet, bFet, and rhFet) exist in very different proteoform populations, as this likely affects their function and should thus be taking into consideration.

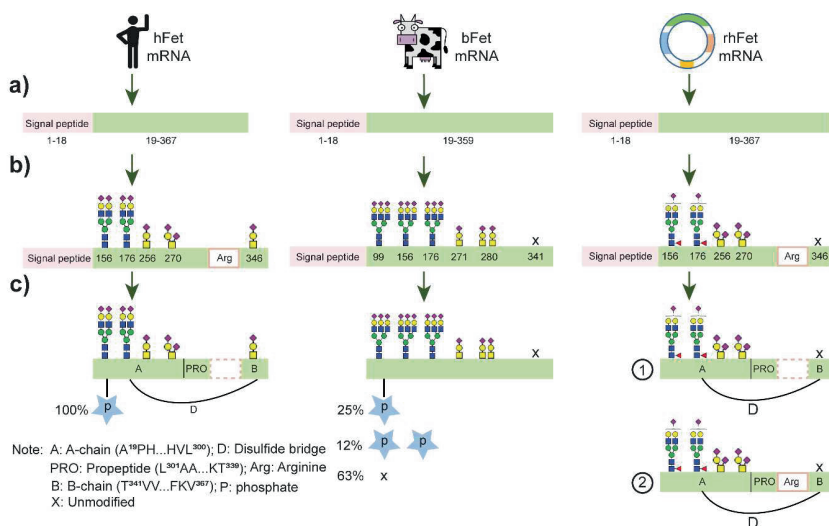


Figure 4. Overall comparison of the protein structure and occurring PTMs on hFet, bFet and rhFet. After (a) translation from single transcript, single-chain preproteins of hFet and bFet contain 367 and 365 amino acids, respectively. Then (b) N-glycosylation occurs in the endoplasmic reticulum followed by the modification of O-glycosylation sites. As in Figure 3 we depict the most abundant proteoforms at each glycosylation site. Distinctively, a high degree of fucosylation occurs on the N-glycosylation sites of rhFet, not observed for hFet. Also, bFet contains mainly triantennary complex glycans, whereas hFet predominantly biantennary glycan structures. The dominant O-glycans in all fetuins are of the core 1 mucin-type harboring one or two sialic acids. (c) The final step, proteolytic processing and phosphorylation likely happen after glycosylation. The signal peptides are removed from the preproteins. For hFet, some unknown proteinase cleaves the C-terminal Arginine at position 322 and converts hFet into a two-chain polypeptide form as we describe in more detail in Figure S3. Interestingly, we identified both the single-chain and two-chain polypeptide form in rhFet, suggesting that in the recombinant expression system the cleavage of Arginine is incomplete (Figure S4). For the phosphorylation, we found all proteoforms of the hFet are fully monophosphorylated; whereas the phosphosite occupancy on bFet was about 25% monophosphorylated, 12% doubly phosphorylated and 63% non-phosphorylated. On rhFet we found no evidence at all for protein phosphorylation.

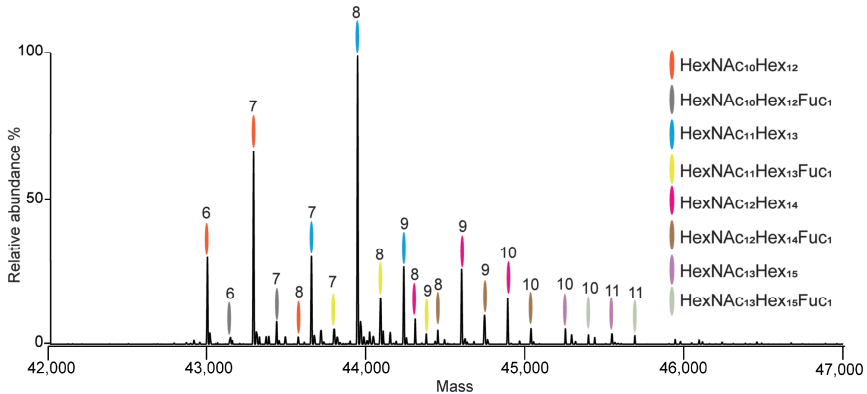


Figure 5. The fully annotated zero charge deconvoluted native mass spectrum of hFet. The overall PTM compositions were assigned based on the accurate mass measurements of the intact protein proteoforms. All proteoforms contain 1 phosphate moiety. The number of sialic acids attached is marked at the top of each peak. For example, the most abundant peak is marked in blue and number 8, as it corresponds to the glycan composition HexNAc₁₁Hex₁₃Neu5Ac₈ and one phosphate moiety.

Serum derived hFet and bFet are well-studied glycoproteins and bear both N- and O-linked glycans. bFet is often used as a model glycoprotein in glycoproteomics. Therefore, we did not expect any surprise observations in our analysis and indeed, our findings are in good agreement with earlier studies. Nonetheless, our hybrid MS approach has the capacity to provide additional information regarding overall structural heterogeneity of the fetuins, which includes not only site-specific characterization of their PTMs, but also analysis of their matured primary structure and its possible variants. Although native hFet is known to be present in serum in the form of

a two polypeptide chain linked by disulfide⁴⁶, no study provided evidence clarifying whether the processing of the hFet primary structure results in one or more sequence variants. We confirm here the existence of the proposed two-chain form architecture of hFet and further show that the cleavage of Arginine at position 322 is in native hFet complete. In our hFet samples, no single-chain proteoforms or proteoforms missing the propeptide were detected. In contrast to the relatively simple proteoform profile of hFet, the native spectrum of rhFet exhibits remarkable more complexity. Major differences between hFet and rhFet originate from more complex glycosylation due to extensive core fucosylated glycans on a various number of antennas. Furthermore, our data revealed that rhFet exists in two sequence variants differing from each other by the absence/presence of the C-terminal Arginine on the A-chain. This is likely caused by incomplete processing of the rhFet in the HEK-293 cells. In consequence, rhFet proteoforms co-exist as a mixture of the one- and two-chain polypeptide form, creating another source of structural diversity. Another striking difference between hFet and rhFet is that the former is for 100% a singly phosphorylated protein, whereas phosphorylation is completely missing in rhFet.

These findings seriously question whether the rhFet studied here represents a good model for wild-type serum hFet. Commercially available recombinant fetuin may be produced by various expression systems and is mostly used for scientific purposes. For example, recombinant fetuin produced by insect cells has been used for studies on the inhibitory effect of human fetuin on insulin-induced autophosphorylation of the insulin receptor^{15,8,14}. Relatively recently, FLAG-tagged human fetuin synthesized in HEK-293T cells has been used to define the mechanism by which fetuin modulates cellular adhesion⁴⁷. With respect to our findings on rhFet, we propose, that any future functional study performed with fetuin produced in HEK-293 cells (or any other expression system) should be critically evaluated, given the distinct structural differences demonstrated here in between hFet and rhFet.

SUPPORTING INFORMATION:

The following supporting information is available free of charge at ACS website <http://pubs.acs.org>

Figure S1 – Amino acid sequence alignment between hFet and bFet

Figure S2 – Native MS spectra of hFet treated by PNGaseF, sialidase and alkaline phosphatase

Figure S3 – A schematic of the two polypeptide chain structure of hFet

Figure S4 – LC-MS/MS spectra of Lys-C peptides from rhFet

Figure S5 – A comparison of the intact protein native MS spectrum with the constructed spectrum based on the peptide-centric proteomics data in hFet and bFet

Table S1 – The list of all modified peptide isoforms on hFet, bFet and rhFet and their relative quantification

Table S2 – The list containing identified hFet and bFet proteoforms

Supplementary Data – Annotated HCD/EThcD spectra of all glycopeptides observed for hFet

Acknowledgments

We acknowledge support from the Netherlands Organization for Scientific Research (NWO) funding the large-scale proteomics facility *Proteins@Work* (project 184.032.201) embedded in the Netherlands Proteomics Centre. A.J.R.H. acknowledge further support by the NWO TOP-Punt Grant 718.015.003. This project has received additional funding from the European Union's Horizon 2020 research and innovation programme under grant agreement 686547 (MSMed).

References

- (1) Bürgi, W.; Schmid, K. Preparation and Properties of Zn-Alpha 2-Glycoprotein of Normal Human Plasma. *J. Biol. Chem.* **1961**, *236* (4), 1066–1074.
- (2) Heremans, J. F. Les Globulines Du Système Gamma Du Plasma Humain; Brussels: Editions Arscia S.A. Paris; **1961**; pp 119-138, DOI: doi.org/10.5169/seals-307477.
- (3) Pedersen, K. O. Fetuin, A New Globulin Isolated from Serum. *Nature* **1944**, *154*, 575.
- (4) Dziegielewska, K.; Brown, W.; Casey, S.; Christie, D.; Foreman, R.; Hill, R.; Saunders, N. The Complete cDNA and Amino Acid Sequence of Bovine Fetuin. *J. Biol. Chem.* **1990**, *265* (6), 4354–4357.
- (5) Brown, W. M.; Dziegielewska, K. M.; Saunders, N. R.; Christie, D. L.; Nawratil, P.; Müller-Esterl, W. The Nucleotide and Deduced Amino Acid Structures of Sheep and Pig Fetuin: Common Structural Features of the Mammalian Fetuin Family. *Eur. J. Biochem.* **1992**, *205* (1), 321–331.
- (6) Schinke, T.; Amendt, C.; Trindl, A.; Pöschke, O.; Müller-Esterl, W.; Jahnen-Dechent, W. The Serum Protein Alpha2-HS Glycoprotein/Fetuin Inhibits Apatite Formation in Vitro and in Mineralizing Calvaria Cells. *J. Biol. Chem.* **1996**, *271* (34), 20789–20796.
- (7) Reynolds, J. L. Multifunctional Roles for Serum Protein Fetuin-A in Inhibition of Human Vascular Smooth Muscle Cell Calcification. *J. Am. Soc. Nephrol.* **2005**, *16* (10), 2920–2930.
- (8) Mathews, S. T.; Chellam, N.; Srinivas, P. R.; Cintron, V. J.; Leon, M. A.; Goustin, A. S.; Grunberger, G. Alpha 2-HSG, a Specific Inhibitor of Insulin Receptor Autophosphorylation, Interacts with the Insulin Receptor. *Mol. Cell. Endocrinol.* **2000**, *164* (1–2), 87–98.
- (9) Janapatla, R. P.; Hsu, M. H.; Liao, W. T.; Chien, K. Y.; Lee, H. Y.; Chiu, C. H. Low Serum Fetuin-A as a Biomarker to Predict Pneumococcal Necrotizing Pneumonia and Hemolytic Uremic Syndrome in Children. *Med. (United States)* **2016**, *95* (13), e3221.
- (10) Laughlin GA, McEvoy LK, Barrett-Connor E, Daniels LB, I. J. Fetuin-A, a New Vascular Biomarker of Cognitive Decline in Older Adults. *Clin Endocrinol* **2014**, *81* (1), 134–140.
- (11) Jensen, M. K.; Bartz, T. M.; Mukamal, K. J.; Djoussé, L.; Kizer, J. R.; Tracy, R. P.; Ziemann, S. J.; Rimm, E. B.; Siscovick, D. S.; Shlipak, M.; et al. Fetuin-A, Type 2 Diabetes, and Risk of Cardiovascular Disease in Older Adults: The Cardiovascular

- Health Study. *Diabetes Care* **2013**, *36* (5), 1222–1228.
- (12) Stefan, N.; Fritsche, A.; Weikert, C.; Boeing, H.; Joost, H.; Schulze, M. B. Plasma Fetuin-A Levels and the Risk of Type 2 Diabetes. *Diabetes* **2008**, *57* (10), 2762–2767.
- (13) Robinson, K. N.; Teran-Garcia, M. From Infancy to Aging: Biological and Behavioral Modifiers of Fetuin-A. *Biochimie* **2016**, *124*, 141–149.
- (14) Srinivas PR, Goustin AS, G. G. Baculoviral Expression of a Natural Inhibitor of the Human Insulin Receptor Tyrosine Kinase. *Biochem Biophys Res Commun* **1995**, *208* (2), 879–885.
- (15) Srinivas PR, Deutsch DD, Mathews ST, Goustin AS, Leon MA, G. G. Recombinant Human Alpha 2-HS Glycoprotein Inhibits Insulin-Stimulated Mitogenic Pathway without Affecting Metabolic Signalling in Chinese Hamster Ovary Cells Overexpressing the Human Insulin Receptor. *Cell. Signal.* **1996**, *8* (8), 567–573.
- (16) Wilson, N. L.; Schulz, B. L.; Karlsson, N. G.; Packer, N. H. Sequential Analysis of N- and O-Linked Glycosylation of 2D-PAGE Separated Glycoproteins. *J. Proteome Res.* **2002**, *1* (6), 521–529.
- (17) Arnold, J. N.; Wormald, M. R.; Sim, R. B.; Rudd, P. M.; Dwek, R. A. The Impact of Glycosylation on the Biological Function and Structure of Human Immunoglobulins. *Annu. Rev. Immunol.* **2007**, *25* (1), 21–50.
- (18) Chalabi, S.; Panico, M.; Sutton-Smith, M.; Haslam, S. M.; Patankar, M. S.; Lattanzio, F. A.; Morris, H. R.; Clark, G. F.; Dell, A. Differential O-Glycosylation of a Conserved Domain Expressed in Murine and Human ZP3. *Biochemistry* **2006**, *45* (2), 637–647.
- (19) Rudd, P. M.; Dwek, R. A. Glycosylation: Heterogeneity and the 3D Structure of Proteins. *Crit. Rev. Biochem. Mol. Biol.* **1997**, *32* (1), 1–100.
- (20) Windwarder, M.; Altmann, F. Site-Specific Analysis of the O-Glycosylation of Bovine Fetuin by Electron-Transfer Dissociation Mass Spectrometry. *J. Proteomics* **2014**, *108*, 258–268.
- (21) Carr, S. A.; Huddleston, M. J.; Bean, M. F. Selective Identification and Differentiation of N- and O-Linked Oligosaccharides in Glycoproteins by Liquid Chromatography-Mass Spectrometry. *Protein Sci.* **1993**, *2* (2), 183–196.
- (22) Kellermann J, Haupt H, Auerswald EA, M.-E. W. The Arrangement of Disulfide Loops in Human Alpha 2-HS Glycoprotein. *J Biol Chem* **1989**, *264* (24), 14121–

14128.

- (23) Spiro, R. G. Protein Glycosylation: Nature, Distribution, Enzymatic Formation, and Disease Implications of Glycopeptide Bonds. *Glycobiology* **2002**, *12* (4), 43R–56R.
- (24) Seo, J.; Lee, K.-J. Post-Translational Modifications and Their Biological Functions: Proteomic Analysis and Systematic Approaches. *J. Biochem. Mol. Biol.* **2004**, *37* (1), 35–44.
- (25) Huang, L. J.; Lin, J. H.; Tsai, J. H.; Chu, Y. Y.; Chen, Y. W.; Chen, S. L.; Chen, S. H. Identification of Protein O-Glycosylation Site and Corresponding Glycans Using Liquid Chromatography-Tandem Mass Spectrometry via Mapping Accurate Mass and Retention Time Shift. *J. Chromatogr. A* **2014**, *1371*, 136–145.
- (26) Haglund, A. C.; Ek, B.; Ek, P. Phosphorylation of Human Plasma Alpha2-Heremans-Schmid Glycoprotein (Human Fetuin) in Vivo. *Biochem J* **2001**, *357* (Pt 2), 437–445.
- (27) Aebi, M. Biochimica et Biophysica Acta N-Linked Protein Glycosylation in the ER ☆. *BBA - Mol. Cell Res.* **2013**, *1833* (11), 2430–2437.
- (28) Dell, A.; Galadari, A.; Sastre, F.; Hitchen, P. Similarities and Differences in the Glycosylation Mechanisms in Prokaryotes and Eukaryotes. *Int. J. Microbiol.* **2010**, *2010*, 148178, DOI:[10.1155/2010/148178](https://doi.org/10.1155/2010/148178).
- (29) Yang, Y.; Liu, F.; Franc, V.; Halim, L. A.; Schellekens, H.; Heck, A. J. R. Hybrid Mass Spectrometry Approaches in Glycoprotein Analysis and Their Usage in Scoring Biosimilarity. *Nat. Commun.* **2016**, *7*, 13397.
- (30) Franc, V.; Yang, Y.; Heck, A. J. R. Proteoform Profile Mapping of the Human Serum Complement Component C9 Revealing Unexpected New Features of N-, O-, and C-Glycosylation. *Anal. Chem.* **2017**, *89* (6), 3483–3491.
- (31) Franc, V.; Zhu, J.; Heck, A. J. R. Comprehensive Proteoform Characterization of Plasma Complement Component C8 $\alpha\beta\gamma$ by Hybrid Mass Spectrometry Approaches. *J. Am. Soc. Mass Spectrom.* **2018**, *29* (6), 1099–1110.
- (32) Rosati, S.; Yang, Y.; Barendregt, A.; Heck, A. J. R. Detailed Mass Analysis of Structural Heterogeneity in Monoclonal Antibodies Using Native Mass Spectrometry. *Nat. Protoc.* **2014**, *9* (4), 967–976.
- (33) Rose, R. J.; Damoc, E.; Denisov, E.; Makarov, A.; Heck, A. J. R. High-Sensitivity Orbitrap Mass Analysis of Intact Macromolecular Assemblies. *Nat Methods* **2012**, *9* (11), 1084–1086.

- (34) Bern, M.; Caval, T.; Kil, Y. J.; Tang, W.; Becker, C.; Carlson, E.; Kletter, D.; Sen, K. I.; Galy, N.; Hagemans, D.; et al. Parsimonious Charge Deconvolution for Native Mass Spectrometry. *J. Proteome Res.* **2018**, *17* (3), 1216–1226.
- (35) Li, J.; Richards, J. C. Functional Glycomics and Glycobiology: An Overview. In *Functional Glycomics: Methods and Protocols*; Li, J., Ed.; Humana Press: Totowa, NJ, 2010; pp 1–8.
- (36) Kussmann, M.; Nordhoff, E.; Rahbek-Nielsen, H.; Haebel, S.; Rossel-Larsen, M.; Jakobsen, L.; Gobom, J.; Mirgorodskaya, E.; Kroll-Kristensen, A.; Palm, L.; et al. Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry Sample Preparation Techniques Designed for Various Peptide and Protein Analytes. *J. Mass Spectrom.* **1997**, *32* (6), 593–601.
- (37) Bern, M.; Kil, Y. J.; Becker, C. Byonic: Advanced Peptide and Protein Identification Software. *Curr. Protoc. Bioinforma.* **2012**, CHAPTER (Unit13.20).
- (38) Wang, C.; Yuan, J.; Wang, Z.; Huang, L. Separation of One-Pot Procedure Released O-Glycans as 1-Phenyl-3-Methyl-5-Pyrazolone Derivatives by Hydrophilic Interaction and Reversed-Phase Liquid Chromatography followed by Identification Using Electrospray Mass Spectrometry and Tandem Mass Spectrometry. *J. Chromatogr. A* **2013**, *1274*, 107–117.
- (39) Hayase, T.; Rice, K. G.; Dziegielewska, K. M.; Kuhlenschmidt, M.; Reilly, T.; Lee, Y. C. Comparison of N-Glycosides of Fetuins from Different Species and Human Alpha 2-HS-Glycoprotein. *Biochemistry* **1992**, *31* (20), 4915–4921.
- (40) Edge, A. S.; Spiro, R. G. Presence of an O-Glycosidically Linked Hexasaccharide in Fetuin. *J. Biol. Chem.* **1987**, *262* (33), 16135–16141.
- (41) Hoffmann, M.; Marx, K.; Reichl, U.; Wuhler, M.; Rapp, E. Site-Specific O - Glycosylation Analysis of Human Blood Plasma Proteins. *Mol. Cell. Proteomics* **2016**, *15* (2), 624–641.
- (42) Halim, A.; Rüetschi, U.; Larson, G.; Nilsson, J. LC-MS/MS Characterization of O-Glycosylation Sites and Glycan Structures of Human Cerebrospinal Fluid Glycoproteins. *J. Proteome Res.* **2013**, *12* (2), 573–584.
- (43) Halim, A.; Nilsson, J.; Rüetschi, U.; Hesse, C.; Larson, G. Human Urinary Glycoproteomics; Attachment Site Specific Analysis of N - and O -Linked Glycosylations by CID and ECD. *Mol. Cell. Proteomics* **2012**, *11* (4), M111.013649.

- (44) Brown WM, Saunders NR, Møllgård K, D. K. Fetuin--an Old Friend Revisited. *Bioessays* **1992**, *14* (11), 749–755.
- (45) Elzanowski, A.; Barker, W. C.; Hunt, L. T.; Seilbawl-Ross, E. Cystatin Domains in Alpha-2-HS-Glycoprotein and Fetuin. *FEBS Lett.* **1988**, *227* (2), 167–170.
- (46) Araki, T.; Yoshioka, Y.; Schmid, K. The Position of the Disulfide Bonds in Human Plasma Alpha 2 HS-Glycoprotein and the Repeating Double Disulfide Bonds in the Domain Structure. *Biochim. Biophys. Acta* **1989**, *994* (3), 195–199.
- (47) Nangami G, Koumangoye R, Shawn Goodwin J, Sakwe AM, Marshall D, Higginbotham J, O. J. Fetuin-A Associates with Histones Intracellularly and Shuttles Them to Exosomes to Promote Focal Adhesion Assembly Resulting in Rapid Adhesion and Spreading in Breast Carcinoma Cells. *Exp Cell Res* **2014**, *328* (2), 388–400.

Supporting information for Chapter 2

Supplementary Figures

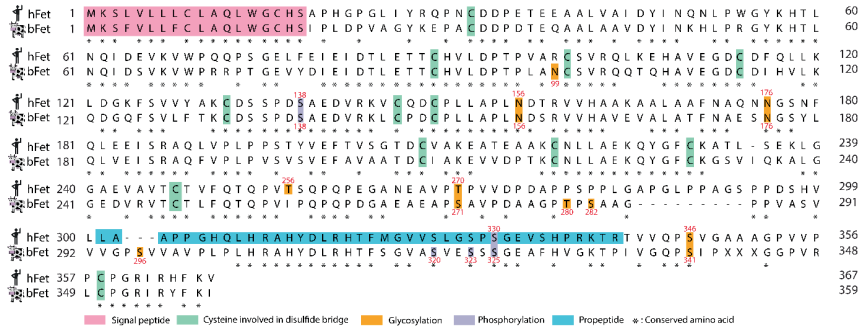


Figure S1. Amino acid sequence alignment between hFet and bFet reveals around 70% sequence similarity. Identical amino acids are marked by ‘*’. Glycosylation and phosphorylation sites are boxed in orange and purple, respectively; the signal peptide is boxed in pink; the propeptide in hFet is boxed in blue. The observed modified amino acids are indicated in red. Both hFet and bFet contain 12 conserved cysteine residues (boxed in light green) involved in 6 disulfide bridges.

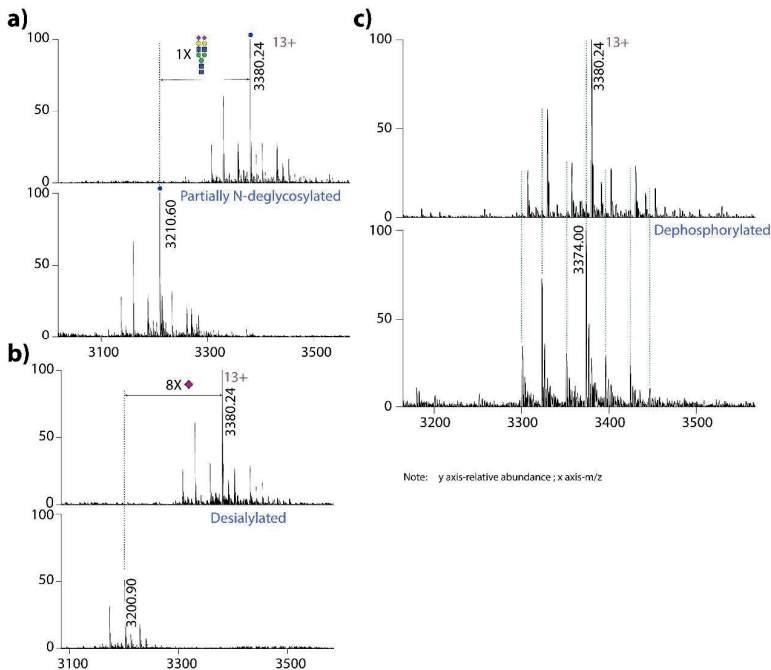


Figure S2. Zoom in on the 13⁺ charge state of hFet to clearly display the changes after specific enzyme treatment. In (a), hFet was treated by PNGase F and revealed partially N-glycosylation in all major peaks; peaks marked with blue color before and after treatment represent the mass difference of one of the N-glycans (~2206 Da). In (b), hFet was treated by sialidase. Comparing the most abundant peak on the charge state 13⁺ with *m/z* of 3380.24 before treatment and 3200.90 after treatment, the mass difference corresponds to 8 sialic acids. In (c), hFet was treated by alkaline phosphatase. The fully phosphorylated modification of hFet was confirmed by the mass shift of 80 Da of all proteoforms after the phosphatase treatment.

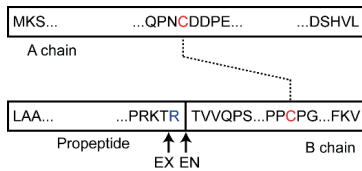


Figure S3. The two polypeptide chain structure of hFet is depicted by a schematic including the A chain at the top, and the connecting peptide and B chain at the bottom. A so far unknown endoproteinase (EN) generates the two-chain form of hFet and an exopeptidase (EX) eliminates the Arginine at position 322, at the end of the propeptide. The A chain and B chain are in hFet connected by a disulfide bridge represented by the dashed line (...).

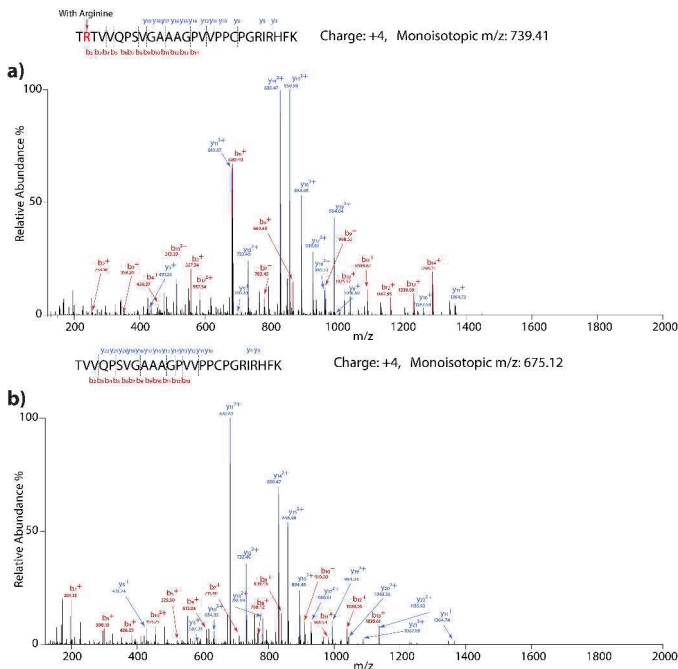


Figure S4. LC-MS/MS spectra of peptides derived from Lys-C/Glu-C digestion of rhFet providing evidence of the co-existence of rhFet proteoforms with and without an Arginine at position 322 (a) a peptide harboring the Arginine and (b) a peptide without the Arginine.

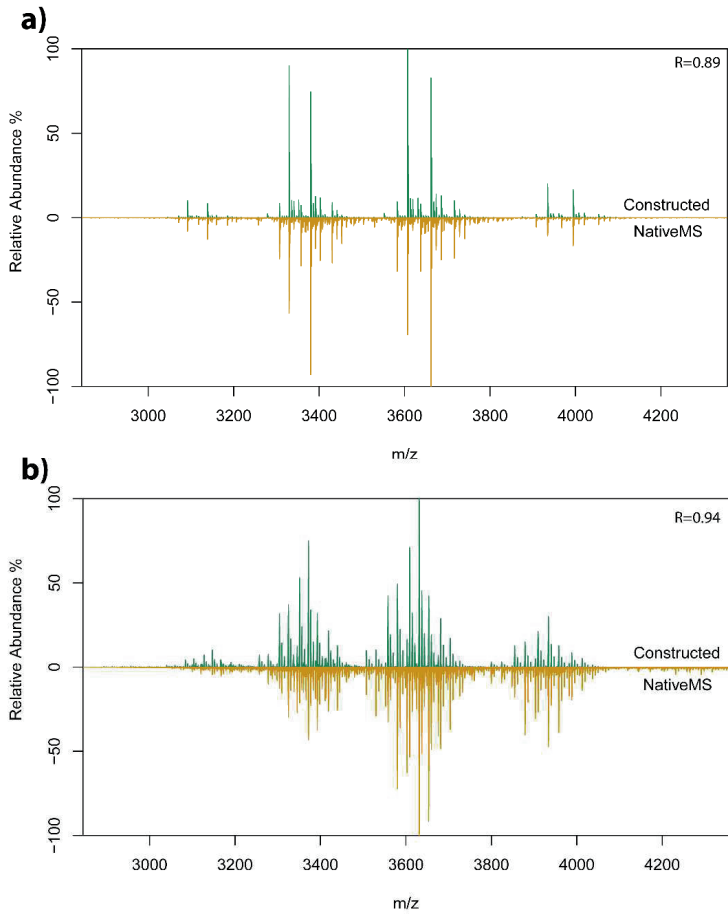


Figure S5. A comparison of the intact protein native MS spectrum with the constructed spectrum based on the peptide-centric proteomics data in **(a)** hFet and **(b)** bFet. The correlation between the spectra is very good, being 0.89 and 0.94 for hFet and bFet, respectively.

The following supplementary tables are not printed because they are oversized.

Supplementary Table S1: The list of all modified peptide isoforms on hFet, bFet and rhFet and their relative quantification

Supplementary Table S2: The list containing identified hFet and bFet proteoforms

Supplementary Data: Annotated HCD/EThcD spectra of all glycopeptides observed for hFet

Chapter 3

Glycoproteogenomics: a frequent gene polymorphism affects the glycosylation pattern of the human serum fetuin/ α -2-HS-glycoprotein

Yu-Hsien Lin^{ab}, Jing Zhu^b, Sander Meijer^{ac}, Vojtech Franc^{ab*}, and Albert J.R. Heck^{ab*}

^aBiomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

^bNetherlands Proteomics Center, Padualaan 8, 3584 CH Utrecht, The Netherlands

^cDepartment of Molecular and Cellular Hemostasis, Sanquin Research, Amsterdam 1066 CX, the Netherlands

* Authors for Correspondence: Vojtech Franc, v.franc@uu.nl (+31302536149) or Albert Heck, a.j.r.heck@uu.nl (+31302536797)

Running title: gene polymorphism affects glycosylation of α -2-HS-glycoprotein

Based on:

Lin, Y. H.; Zhu, J.; Meijer, S.; Franc, V.; Heck, A. J. R. Glycoproteogenomics: A Frequent Gene Polymorphism Affects the Glycosylation Pattern of the Human Serum Fetuin/ α -2-HS-Glycoprotein. *Mol. Cell. Proteomics*. 2019, 18 (8), 1479-1490.

Abbreviations

ABC

ammonium bicarbonate

AMAC

ammonium acetate

ACN

Acetonitrile

Asn

Asparagine

CAT

weak cation exchange

CKD

chronic kidney disease

CVD

cardiovascular disease

dHex

deoxyhexose

EMR

Extended Plus Orbitrap instrument

ELISA

enzyme-linked immunosorbent assays

ESI

electrospray ionization

ETheD

electron-transfer combined with higher-energy collision dissociation

FA

formic acid

GlcNAc

N-acetylglucosamine

HexNAc

N-acetylhexosamine

IAA

iodoacetamide

IEX

ion-exchange chromatography

INSR

insulin receptor

Met

Methionine

MS

mass spectrometry

Neu5Ac

N-acetylneuraminic acid

Pho

phosphorylation

PTMs

posttranslational modifications

Ser

Serine

SNPs

single nucleotide polymorphisms

T2DM

type 2 diabetes

Thr

Threonine

XIC

extracted ion chromatogram

WAX

weak anion exchange

Abstract

Fetuin, also known as α -2-HS-glycoprotein (gene name: AHSB), is one of the more abundant glycoproteins secreted into the bloodstream. There are two frequently occurring alleles of human AHSB, resulting in three genotypes (AHSB*1, AHSB*2, and heterozygous AHSB1/2). The backbone amino acid sequences of fetuin coded by the AHSB*1 and AHSB*2 genes differ in two amino acids including one known O-glycosylation site (aa position 256). Although fetuin levels have been extensively studied, the originating genotype is often ignored in such analysis. As fetuin has been suggested repeatedly as a potential biomarker for several disorders, the question whether the gene polymorphism affects the fetuin profile is of great interest. In this work, we describe detailed proteoform profiles of fetuin, isolated from serum of 10 healthy and 10 septic patient individuals and investigate potential glycoproteogenomics correlations, e.g. how gene polymorphisms affect glycosylation. We established an efficient method for fetuin purification from individuals' serum using ion-exchange chromatography. Subsequently, we performed hybrid mass spectrometric approaches integrating data from native mass spectra and peptide-centric MS analysis. Our data reveal a crucial effect of the gene polymorphism on the glycosylation pattern of fetuin. Moreover, we clearly observed increased fucosylation in the samples derived from the septic patients. Our serum proteoform analysis, targeted at one protein obtained from 20 individuals, exposes the wide variability in proteoform profiles, which should be taken into consideration when using fetuin as biomarker. Importantly, focusing on a single or few proteins, the quantitative proteoform profiles can provide, as shown here, already ample data to classify individuals by genotype and disease state.

Introduction

Human fetuin, also known as α -2-HS glycoprotein, is an abundant glycoprotein circulating in human plasma. In adults, fetuin is secreted almost exclusively by the liver (1), but many other cells are able to produce this protein (2, 3). Since its discovery in 1944, human fetuin has been ascribed to various biological functions, but a concrete understanding of its exact role remains poor. One widely accepted function is its role in calcium and phosphate metabolism. Fetuin acts as one of the essential inhibitors in the prevention of ectopic mineral precipitation and deposition (4–8). Additionally, several studies have suggested a correlation between circulating fetuin levels and inflammatory, and chronic diseases, such as endotoxemia and sepsis (9), cardiovascular disease (CVD), type 2 diabetes (T2DM) (10, 11) and chronic kidney disease (CKD) (12), raising interest in its use as a protein biomarker.

Values of serum or plasma fetuin in healthy humans range from 300 to 750 μ g/ml (13, 14). Fetuin concentrations seem to be independent of gender, but serum fetuin levels correlate with age and might also be influenced by genetic background (15, 16). A highly frequent genetic polymorphism of the fetuin gene (AHSG), resulting from the two common alleles AHSG*1 and AHSG*2, was already described three decades ago (17). Genomic analysis has shown that the AHSG*1 allele differs from AHSG*2 by two amino acids at positions 248 and 256, respectively. The AHSG*1 allele is characterized by ACG (Thr248) and ACC (Thr256), whereas the AHSG*2 allele is characterized by ATG (Met248) and AGC (Ser256) (18). These two abundant alleles of AHSG result in three common genotypes (AHSG*1, AHSG*2, and heterozygous AHSG1/2). The population distribution of these three genotypes varies globally. The available data from several surveys demonstrate that AHSG*1 is in frequency the foremost fetuin genotype (19–21). For example, the distribution of AHSG*1, AHSG1/2 and AHSG*2 in 697 unrelated Italians was found to be around 0.56:0.36:0.08 (22). Some studies have attempted to investigate a potential correlation between this fetuin gene polymorphism and certain diseases, however, the reported data have not been very conclusive (10, 16).

Human fetuin is a glycoprotein and its primary structure and glycosylation profile have been described (23, 24). However, the primary structure of fetuin is somehow elusive and reported discrepancies lead to confusion especially related to the absence/presence of propeptide

(connecting peptide) in the matured human fetuin. According to our recent work (25), the primary structure of human fetuin contains the connecting peptide and is different from well-studied bovine fetuin. The matured human fetuin comprises an A-chain with the connecting peptide, connected to a smaller B-chain *via* a single disulfide linkage. Human fetuin is modified with a variety of posttranslational modifications (PTMs); it harbors several N- and O-glycosylations and a few phosphorylations, generating a complex mixture of proteoforms. In more detail, human fetuin harbors two N-glycosylation sites and two O-glycosylation sites in the A-chain, and one O-glycosylation site in the B-chain (25, 26). Interestingly, one A-chain O-glycan is located at position Thr256 in the AHSG*1 genotype, which is exactly the site involved in the most frequent gene polymorphism. Whether the replacement of the Thr by the Ser at position 256 in AHSG*2 affects the glycosylation pattern of fetuin has yet not been addressed. In this particular case, the Thr to Ser exchange still allows either amino residue to be O-glycosylated. In general, PTMs can modulate protein structure and play a key role in regulating the stability and physical-chemical properties of proteins (27, 28) and therefore there is a need to study proteoform profiles in detail. Human fetuin has been considered as a biomarker for certain metabolic diseases, but its quantitative analysis by various commercial enzyme-linked immunosorbent assays (ELISA) has been found to be problematic (29), which we hypothesize may be partly related to differences in the glycosylation patterns of AHSG*1 and AHSG*2.

Since fetuin is also known to be a negative acute phase reactant, the relationship between fetuin levels in plasma and the immune response is also of considerable interest. Glycosylation patterns of acute phase proteins may change in response to diseases such as cancer (30) and during acute inflammation. Consequently, aberrant glycosylation patterns have been related to the development or progression of disease (31–33). For example, the elevated expression of sialyl-lewis X antigen has been observed on haptoglobin, α -1-antichymotrypsin, and α -1-acid glycoprotein in response to sepsis (34, 35). Although site-specific N-glycan profiles of these acute phase proteins have been obtained before, there has been, as far as we know, no study

dealing with glycosylation changes possibly occurring in fetuin during acute and chronic inflammation.

Here, we aim to describe proteoform profiles of fetuin directly isolated from serum of healthy and septic patient individuals and fill in these knowledge gaps, thereby especially focusing on glycoproteogenomics correlations, e.g. how gene polymorphisms may affect glycosylation profiles. This represents a challenging task, not only due to the complicated proteoform profile of fetuin, but also due to the fact that fetuin has to be purified from human serum, which contains several more abundant (glyco)proteins. We demonstrated recently that using state-of-the-art hybrid mass spectrometry (MS) approaches, combining high-resolution native MS and peptide-centric MS, we may comprehensively characterize all PTMs on various therapeutic and serum glycoproteins (36–38). Making use of such methods, we previously also reported on the very distinct proteoform profiles of fetuin obtained from three different biological sources, namely plasma-derived human fetuin from a large pool of donors, recombinant human expressed in HEK-293 cells and bovine fetuin (25). Here, we first set out to extend that latter study, focusing on fetuin extracted from individuals. We therefore first purified endogenous fetuin from several individuals' serum (healthy and septic patient donors) using a one-step mixed-mode ion-exchange chromatography fractionation (IEX) (39) and performed subsequently an in-depth characterization of the proteoform profiles. We introduce an algorithm enabling semi-automated assignment of the native MS spectra based on the data gathered from the glycopeptide-centric measurements. Our data reveal, for the first time, a dramatic effect of the gene polymorphism on the glycosylation pattern of fetuin, which needs to be taken into account when fetuin would be used as a biomarker. Based on this knowledge we are able to obtain and annotate the full proteoform profiles of fetuin from 20 individuals. These proteoform profiles allow classification into the three different genotypes. Concomitantly, looking at fetuin purified from 10 healthy individual donors and 10 septic patient donors we observe significantly increased fucosylation, especially at site N176, in the samples derived from the septic patients. Our serum fetuin analysis clearly exposes the wide

variability in individual proteoform profiles. However, we demonstrate that these fetuin proteoform profiles may provide data that can be used to classify individuals by genotype and disease state.

Experimental procedures:

Chemicals

Dithiothreitol (DTT), iodoacetamide (IAA), trifluoroacetic acid (TFA), formic acid (FA), ammonium bicarbonate (ABC) and ammonium acetate (AMAC) were purchased from Sigma-Aldrich (Steinheim, Germany). The fetuin sample used as reference standard was purified from a pool of human serum samples and as such acquired from Sigma (α -2-HS glycoprotein; Uniprot Code: P02765, Sigma-Aldrich, Steinheim, Germany). Acetonitrile (ACN) was purchased from Biosolve (Valkenswaard, The Netherlands). Sequencing grade trypsin was obtained from Promega (Madison, WI). Gluc-C was obtained from Roche (Indianapolis, USA).

Individual serum samples

Individual human serum from 10 anonymous healthy donors was provided by Sanquin Research, Amsterdam, The Netherlands. Plasma samples were obtained after informed consent and in accordance to the ethics board of Sanquin. The whole blood from each individual was collected in a 9-mL Vacuette tube (Greiner Bio-One, Kremsmunster, Austria) containing Z Serum Clot Activator. The whole blood was then left undistributed at room temperature for 30-60 min. The clotted material was removed by centrifugation at 1800 x g for 20 minutes at room temperature and the sera were transferred as 1 mL aliquots to clean 1.5 mL Eppendorf tubes, snap froze in liquid nitrogen and stored at -80 °C. Human serum samples from 10 septic patients were acquired from Discovery Life Science, Columbus, OH, USA. All serum samples were stored at -80°C until analysis.

Experimental design and statistical rationale

We purified fetuin and conducted native MS and peptide-centric MS analysis of serum fetuin obtained from 20 individuals (10 healthy donors and 10 septic patient donors). Our mass spectrometric data suggested that each fetuin genotype contains unique features. These features, present in all 20 serum samples, were used for classification (using a Pearson correlation coefficient) of the donors into their genotype. Following genotypic classification, we looked at the differences in fucosylation between healthy and septic donors. For that, we applied a standard T-test to determine if there is a statistically significant difference between these two groups of donors.

Dual-column ion-exchange chromatography for fetuin purification from individual serum samples

An Agilent 1290 Infinity HPLC system (Agilent Technologies, Waldbronn, Germany) consisting of a vacuum degasser, binary pump, refrigerated autosampler with 500- μ L injector loop, two column compartment with thermostat, auto collection fraction module and multi-wavelength detector, was used in this study. The dual-column set-up, comprised of a tandem WAX-CAT (PolyWAX LP, 200 \times 2.1 mm i.d., 5 μ m, 1000 Å; PolyCAT A, 50 \times 2.1 mm i.d., 5 μ m, 1000 Å) two-stage column set-up. All columns were obtained from PolyLC Inc. (Columbia, USA) (39). The column compartment was cooled to 17 °C while the other compartments were chilled to 4 °C to minimize sample degradation. Mobile phase Buffer A consisted of 100 mM AMAC in water, and mobile phase Buffer B consisted of 2.5 M AMAC in water. Typically, 50 μ L of serum sample was injected per run (~3.5 mg total protein). Elution was achieved using a multi-step gradient, consisting of six transitions with increasing proportions of Buffer B: (step 1; equilibration) 0%B, 0–6 min; (step 2; salt gradient) 0–20%B, 6–11 min; (step 3; salt gradient) 20–36%B, 11–24 min; (step 4; high salt rinse) 36–100%B, 24–28 min; (step 5; high salt wash) 100%B, 28–32 min; (step 6; restoration) 100–0%B. The flow

rate was set to 800 $\mu\text{L}/\text{min}$. The chromatograms were monitored by absorption at 280 nm and time-based fractions collected every 0.5 min from 19-23 min using an automated fraction collector.

Native MS analysis of fetuin

The human fetuin fraction, well separated from the most abundant serum proteins IgG and albumin, was collected from in between 22.5-23 min (see Supplemental Figure S1). This fraction, which contained about 25-30 μg of the fetuin protein, was buffer exchanged into 150 mM AMAC (pH 7.2) by ultrafiltration (vivaspin500, Sartorius Stedim Biotech, Germany) with a 10 kDa cut-off filter. The protein concentration was measured by absorbance at 280 nm and adjusted to 2-3 μM prior to native MS analysis.

Collected fetuin fraction was analyzed on a modified Exactive Plus Orbitrap instrument with extended mass range (EMR) (Thermo Fisher Scientific, Bremen) using a standard m/z range of 500-10,000, as described in detail previously (40, 41). The voltage offsets on the transport multi-poles and ion lenses were manually tuned to achieve optimal transmission of protein ions at elevated m/z . Nitrogen was used in the HCD cell at a gas pressure of $6-8 \times 10^{-10}$ bar. The MS parameters were used typically: spray voltage 1.2-1.3 V, source fragmentation 30 V, source temperature 250 $^{\circ}\text{C}$, collision energy 30 V, and resolution (at m/z 200) 17,500. The mass spectrometer was calibrated using CsI clusters as described previously (40).

The accurate masses of observed proteoforms of fetuins were extracted by deconvoluting the native mass spectrum to zero-charge using the Intact Mass software (Protein Metrics ver. 1.5) (42). For the analysis of the PTMs, the data was processed manually, and glycan structures were retrieved on the basis of known biosynthetic pathways. The average masses were used for these calculations; hexose/mannose/galactose (Hex/Man/Gal, 162.1424 Da), N-acetylhexosamine/N-acetylglucosamine (HexNAc/GlcNAc, 203.1950 Da), deoxyhexose (dHex, 146.1430 Da), N-acetylneuraminic acid (Neu5Ac, 291.2579 Da and phosphorylation

(Pho 79.9799 Da). All used symbols and nomenclature are based on the recommendations of the Consortium for Functional Glycomics (43).

In-solution digestion for peptide-centric proteomics

Around 2 µg of fetuin protein was taken from the collected fraction and dissolved into 50 mM ABC at a concentration of 1 mg/mL. The sample was then reduced with 4 mM DTT at 56 °C for 30 min and alkylated with 8 mM IAA at room temperature for 30 min in the dark. Fetuin was digested for 3 h with Glu-C at an enzyme-to-protein-ratio of 1:75 (w/w) at 37 °C and the resulted peptide mixtures were further digested by using trypsin (1:100; w/w). All proteolytic digests containing modified glycopeptides were desalted using Oasis HLB C18 cartridges, dried, and reconstituted in 20 µL of 0.1% FA prior to liquid chromatography (LC)-MS and MS/MS analysis.

LC-MS and MS/MS analysis

All peptides (300 fmol of fetuin peptides) were separated and analyzed using an Agilent 1290 Infinity HPLC system (Agilent Technologies, Waldbronn, Germany) coupled on-line to an Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). Reversed-phase separation was accomplished using a 100 µm inner diameter 2 cm trap column (in-house packed with ReproSil-Pur C18-AQ, 3 µm) (Dr. Maisch GmbH, Ammerbuch-Entringen, Germany) coupled to a 50 µm inner diameter 50 cm analytical column (in-house packed with Poroshell 120 EC-C18, 2.7 µm) (Agilent Technologies, Amstelveen, The Netherlands). Mobile-phase solvent A consisted of 0.1% formic acid in water, and mobile-phase solvent B consisted of 0.1% formic acid in ACN. The flow rate was set to 300 nL/min. A 45 min gradient was used as follows: 0-5 min, 100% solvent A; 13-44% solvent B within 20 min; 44-100% solvent B within 3 min; 100% solvent B for 1 min; 100% solvent A for 17 min. For the MS scan, the mass range was set from 375-1500 *m/z* at a resolution of 120,000 and the

AGC target was to 4×10^5 . For the MS/MS measurements, electron-transfer combined with higher-energy collision dissociation (EThcD) (44) was used and performed with a normalized collision energy of 35%. For the MS/MS scan, the mass range was set from 125-2000 m/z ; the AGC target was set to 5×10^4 . The precursor isolation width was 1.6 Da, and the maximum injection time was set to 200 ms.

LC-MS and MS/MS data analysis

The raw data files containing MS/MS spectra of fetuin peptides were processed using Byonic software (ver 3.2.0) (Protein Metrics Inc.) (45). The following parameters were used for data searches in Byonic: precursor ion mass tolerance, 10 ppm; product ion mass tolerance, 20 ppm; fixed modification, Cys carbamidomethyl; variable modification: Met oxidation, Ser, Thr and Tyr phosphorylation, and both N- and O- glycosylation from mammalian glycan databases; the allowed number of miss-cleavages was set to 4. Trypsin (C-terminal RK) and Glu-C (C-terminal DE) enzyme specificity search was chosen for all samples. The fasta file used for the peptide searches contained both the AHSG*1 and AHSG*2 fetuin amino acid sequences (UniProtKB - P02765, FETUA_HUMAN, AHSG*1 has Thr-248/Thr-256; AHSG*2 has Met-248/Ser-256). Byonic peptide cut-off score of 200 was used and all PTM-modified identified spectra were further manually inspected. Site-specific quantification of the fetuin PTMs was performed as follows; the first three isotopes were taken from each manually validated peptide proteoform for the calculation of the peak areas. Each peptide that contained modified sites was normalized individually so that the sum of all its proteoform areas was set to 100%. The average peptide ratios from all measurements were taken as a final estimation of the abundance. The XICs were obtained using the software Skyline (ver 4.2.0.18305) (46). The glycan structures of each glycoform were manually annotated. Hereby, reported glycan structures are depicted without the linkage type of the glycan units since our acquired MS/MS data do not directly provide such information.

Automated site-specific proteoform annotation in native MS spectra using peptide-centric data

We annotated in a site-specific manner each proteoform represented by peaks in the native MS spectra using the relative abundance of all PTMs sites derived from the peptide-centric data (the algorithm is made publicly available and supplemented at <https://github.com/juer120/NativePTMannotation>). The process of annotation involved two major steps: (1) the relative abundances of all possible PTM combinations on the proteoforms are retrieved from the peptide-centric data; (2) the masses of all proteoforms are extracted from the native MS spectra and then matched with the calculated proteoforms consisting of the PTM combination with the highest relative abundances, and mass window centered around detected proteoform within 100 ppm.

In detail, we first input the mass of the protein backbone retrieved from the protein sequence corrected for the presence of disulfide bridges (F_{pb}) and then calculate the theoretical m/z of all proteoforms in determined charge state (z) by permuting masses of all PTMs identified from peptide-centric measurements. The m/z of a proteoform F^{z+} can be calculated as:

$$F^{z+} = (F_{pb} + \sum_{i=1}^n f_{ij} + z * 1.007276) / z$$

where n is the number of PTM sites; f_{ij} is the mass of the j -th modification at site i . Second, we calculate the relative abundance P of a proteoform F^{z+} with all possible PTM combinations, which is calculated by:

$$P = \prod_{i=1}^n P_{ij}$$

where P_{ij} is the normalized relative abundance of the j -th modification at site i ; n is the number of PTM sites. P_{ij} is calculated by:

$$P_{ij} = \frac{V_{ij}}{\sum_{j=1}^k V_{ij}}$$

where k is the number of possible PTM isoforms at site i (including unoccupied site); V_{ij} is the abundance of j -th modification at site i , which is calculated based on the XIC. Third, we input all m/z of proteoforms (M^{z+}) extracted from raw native MS spectrum and then match them with theoretical m/z (F^{z+}) within 100 ppm, producing a list of all matched m/z . Lastly, we export the highest relative abundance of PTMs combination corresponding to the matched m/z in the list.

Table.1 Overview of donor characteristics. The first column lists the used sample names.

Sample name	Sex	Age	Gene type	Sample type
M1	M	48	AHSG*1	Healthy
F1	F	37	AHSG*1	Healthy
F2	F	26	AHSG*1	Healthy
F5	F	63	AHSG*1	Healthy
M2	M	55	AHSG*2	Healthy
M3	M	54	AHSG*2	Healthy
F3	F	33	AHSG1/2	Healthy
F4	F	27	AHSG1/2	Healthy
M4	M	27	AHSG1/2	Healthy
M5	M	63	AHSG1/2	Healthy
F6	F	68	AHSG*1	Sepsis
F7	F	19	AHSG*1	Sepsis
F8	F	73	AHSG*1	Sepsis
M7	M	18	AHSG*1	Sepsis
F9	F	33	AHSG*1	Sepsis
F10	F	82	AHSG*1	Sepsis
M8	M	47	AHSG1/2	Sepsis
F11	F	55	AHSG1/2	Sepsis
F12	F	33	AHSG1/2	Sepsis
F13	F	37	AHSG1/2	Sepsis

Results

Native mass spectra and peptide-centric analysis of fetuin purified from healthy individuals

We initiated our study characterizing in detail human serum fetuin proteoform profiles of 10 randomly selected healthy individuals (Table 1). We soon found out that these individuals harbor different fetuin genotypes and started to investigate the effect of this genetic polymorphism on the fetuin glycosylation pattern. Since fetuin is produced by at least two frequently occurring variants of the gene AHSG, AHSG*1, and AHSG*2, one could expect to

observe distinct proteoform profiles of fetuin isolated from these individuals. To investigate this, we first loaded 50 μ L of human serum of each individual on the dual mixed-mode column

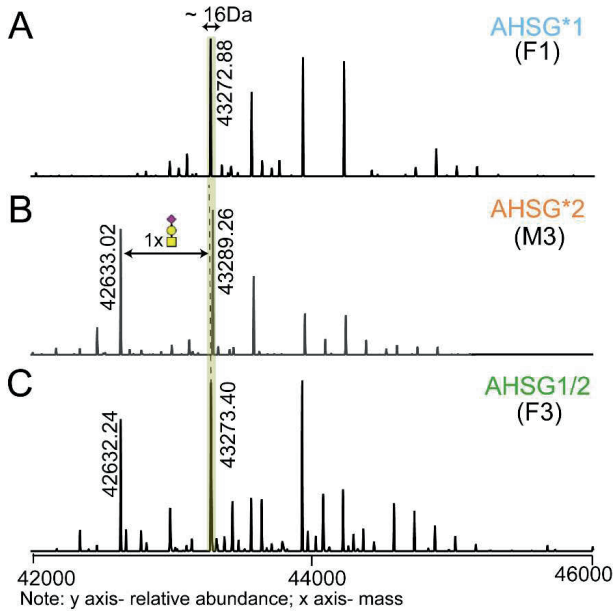


Figure 1. Different genotypes lead to different proteoform profiles. Three representative zero-charge deconvoluted native mass spectra of fetuin purified from serum of A. donor F1 (AHSG*1), B. donor M3 (AHSG*2), and C. donor F3 (AHSG1/2). Corresponding proteoforms between AHSG*1 and AHSG*2 differ in mass from each other by 16 Da (as shown for the proteoforms with the mass of 43,272.88 Da in A and 43,289.26 Da in B). In B, the proteoform with the mass of 42,633.02 Da differs from the most abundant proteoform by 656 Da, which agrees to an O-glycan composition of HexNAc₁Hex₁Neu5Ac₁. This proteoform (42,632.24 Da) can be observed in C, but is fully absent in A.

ion exchange chromatography system, comprising a weak cation exchange (CAT) column preceded by a weak anion exchange (WAX) column. Using this chromatographic set-up for serum, we collected in each run a unique fraction containing fetuin, which was subsequently concentrated, buffer exchanged, and subjected to high-resolution native MS (Supplemental Figure S2). Figure 1 displays three representative zero-charge deconvoluted native mass spectra originating from individuals with the three different genotypes of fetuin (donor F1 - AHSG*1, donor M3 - AHSG*2, and donor F3 - AHSG*1/2). The grouping of all acquired spectra was

done based on signature peaks specific for each genotype, as explained in more detail below. Although all three native mass spectra appeared to be quite distinct, many of the peaks were present in each spectrum and differed from each other primarily in signal intensity. Nevertheless, a closer look into the spectra revealed some key differences. One intense peak with the mass of 42,633.02 Da detected in AHSG*2 and AHSG*1/2 (42,632.24 Da) was completely missing in all AHSG*1 (Figure 1A, and Supplemental Figure S2A). Therefore, this peak was specific for AHSG*2 and AHSG*1/2, and its complete absence defined unambiguously AHSG*1. Interestingly, the missing proteoform in AHSG*1 seemed to represent a significant difference in the proteoform profile between AHSG*1 and AHSG*2 and hinted at a substantial influence of the genetic polymorphism on the glycosylation profile. Additionally, when comparing corresponding peaks between AHSG*1 and AHSG*2, the mass difference of 16 Da could be observed (Figure 1A and B, Supplemental Figure S2A and B). This is due to the mass difference between the backbone amino acid sequences of AHSG*1 and AHSG*2, exactly 15.99 Da. The native mass spectra of fetuin from the heterozygote genotype AHSG*1/2 donors showed more complex proteoform profiles (Figure 1C, Supplemental Figure S2C) compared to the other two genotypes. This is expected, as it should be a combination of all the proteoform profiles observed for AHSG*1 and AHSG*2. In theory, we should detect double peaks for all proteoforms that are common for AHSG*1 and AHSG*2. Nevertheless, due to the very small m/z difference between both AHSG genes (16 Da, i.e. 1.23 Th for the 13⁺ charge state), we observed in practice broader or partially split peaks in the native mass spectra obtained for the heterozygote donors. Resolving such a small mass difference would require approximately 10 times higher mass resolving power in the mass region around 4000 m/z . Our findings from the native mass analysis could be further supported and explained with the subsequent (glyco)peptide-centric analysis of all individual donors' fetuins. We digested fetuins from all three genotypes with trypsin and Gluc-C, which resulted in a set of peptides for subsequent sequencing by LC-MS/MS analysis. Data interpretation provided information about the site location, PTM type, composition and abundance of all identified peptides (for manually annotated MS/MS spectra see Supplemental File S1 and S2; for Byonic searches and MS/MS

spectra see “Data availability”). Figure 2AB and Supplemental File S1 displays annotated EThcD MS/MS spectra of two peptides with amino acid sequence $^{243}\text{VAVTCTIVFQTQPVTSQPQPE}^{262}$ and $^{243}\text{VAVTCMVFQTQPVSSQPQPE}^{262}$, respectively signature peptides derived from proteolytic digestion of AHSG*1 and AHSG*2. These peptides contain the two aforementioned genotypic sites. Notably, manual inspection of the XICs obtained from all three fetuin genotypes revealed that Thr256 on the AHSG*1 peptide was fully occupied with O-glycans containing 0, 1 or 2 sialic acids connected to the core structure HexNAc₁Hex₁, whereas Ser256 on the AHSG*2 peptide was found mostly to be non-modified (Figure 2C, Supplemental Figure S3). Thus, although Ser256 was marginally modified by O-glycan, it differs significantly from Thr256 which was found to be fully O-glycosylated. The relative abundance of all detected variants of the peptides containing these sites across the three different genotypes is displayed in Figure 2C. The relative quantification of the site occupancy of Thr/Ser256, as obtained for all 10 studied healthy individuals, was processed using Skyline and provided in Supplemental Table S1. From the peptide-centric data, we could conclusively explain the origin of the mass difference between the peaks 43,272.88 Da in AHSG*1 and 42,633.02 Da in AHSG*2 (HexNAc₁Hex₁Neu5Ac₁ - 15.99).

Site-specific PTMs annotation of the native mass spectra

We introduced earlier an algorithm facilitating the integration of the data from native mass spectrometry and peptide-centric LC-MS/MS data, which allowed us to assess the integrity of the glycopeptide characterization through *in silico* construction of an intact proteoform profile from all combined peptide-centric data (36). The algorithm lists all possible combinations of proteoform masses based on observed PTMs, at the peptide-centric level, and eventually constructs simulated native-like mass spectra, which can be correlated to the experimental native mass spectra. However, in the earlier approach the resulted list containing all predicted proteoforms, with the total PTM composition, did not contain any site-specific information on the PTMs. Here, we further extended our algorithm (made publicly available) and applied it for

the semi-automated annotation of PTMs to all the peaks in the native mass spectra, now also including site location. To perform this more advanced assignment, we amended our in-house scripts written in R (see “Methods section for the detailed description”). In short, we first generated a library of proteoforms with specific masses and probability ranks using the data (masses and relative intensities) derived from the peptide-centric analysis.

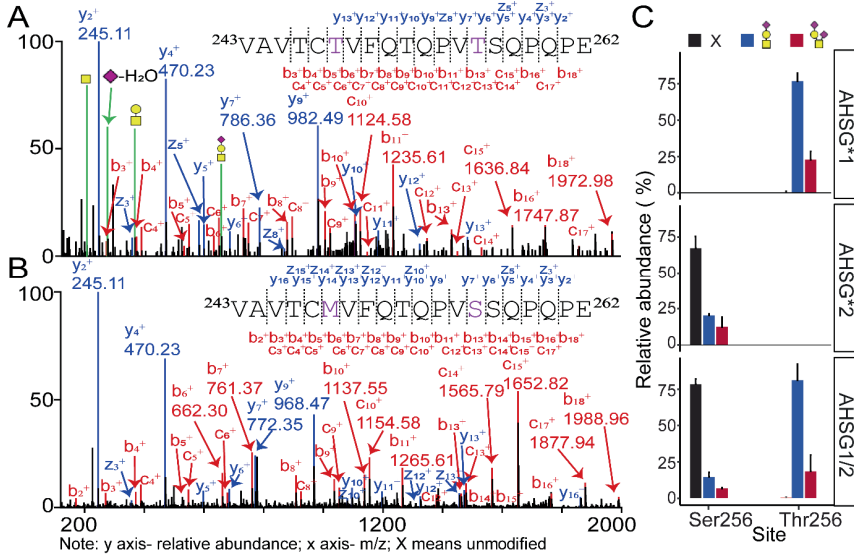


Figure 2. Peptide signatures of distinct fetuin genotypes. ETHcD MS/MS spectra of the peptide A. $^{243}\text{VAVTCTVFQTPVTSQPQPE}^{262}$ (AHSG*1) and B. $^{243}\text{VAVTCTVFQTPVSSQPQPE}^{262}$ (AHSG*2), obtained by proteolytic digestion (Trypsin + Glu-C), harboring the two mutation sites. The ETHcD spectra provide confirmation of the unique peptide sequence and position of the O-glycan, being Thr256 in AHSG*1. Spectra were acquired for the precursor ions with three charges and m/z of 958.78, and 745.03 for AHSG*1 and AHSG*2, respectively as these peptides were predominantly non-modified for AHSG*2 and fully modified for AHSG*1. C. Quantification of the peptide signatures, containing the mutations and O-glycosylation site Thr/Ser256. In these bars, the abundance of these peptides was averaged over all genotype-specific fetuin samples purified from the serum of the 10 healthy individuals across all three genotypes, AHSG*1 (n=4), AHSG*2 (n=2) and AHSG1/2 (n=4). Relative abundances of peptide proteoforms were estimated from their corresponding extracted ion chromatograms (XICs) and normalized to 100%. The AHSG*1 and AHSG*2 signature peptides contain the Thr256 and Ser256 site, respectively. Both of these peptides can be extracted and separately quantified for the heterozygote AHSG1/2 donors. Black bar is unmodified, blue bar is modified by HexNAc₁Hex₁ with one sialic acid and red bar is modified by HexNAc₁Hex₁ harboring two sialic acids.

All proteoform mass combinations were ranked according to the intensity of the detected PTM peptides from which they were constructed and the match between their theoretical and experimental masses. This means that proteoforms with the same masses (within 100 ppm), but different peptide combinations, were sorted based on the better mass match and intensity of the peptides (meaning that closer mass matches and higher intensities do get higher ranks). Next, we discarded all peaks with a relative intensity of less than 5% (related to the base peak) in the native mass spectra. The masses of peaks were then matched to the generated library with the highest ranked proteoforms. This allowed us to visualize the most probable proteoforms in the native mass spectra with their PTM composition in a site-specific manner. Figure 3 shows such a visualization for fetuin which was purified from a pool of donors of all different genotypes (our reference sample). Cumulatively, based on our experimental data, we defined on fetuin one fixed modification (Ser138 phosphorylation) and five other PTM sites (A=Asn156, B=Asn176, C=Thr/Ser256, D=Thr270, and E=Ser346). The script operated with the fetuin backbone mass corrected for the mass shift induced by the disulfide bonds (-12×1.0079 Da) and one phosphate (+79.97) moiety (Ser138). Peaks in the native mass spectra were then automatically annotated based on the library containing the PTM combinations with the highest rank. The final result contained a list of proteoforms with their masses, the most probable PTM combinations and site-specific information as shown in Supplemental Table S2.

Site-specific annotation fully explains differences observed among AHSG*1, AHSG*2 and AHSG*1/2 fetuin

The native mass spectra obtained from genotype-representative individuals with site-specifically annotated proteoforms defining each genotype (i.e. F1/AHSG*1, M3/AHSG*2, and F3/AHSG*1/2) are provided in Supplemental Figure S4. Some major differences observed across all individuals defining fetuin genotypes were already indicated above. However, the exact differences among them could be only clarified after site-specific characterization of all proteoforms. The most striking one represented the high-intensity peak with a mass of

42,633.02 Da, which is completely missing in fetuin from AHSG*1 donors and its composition consisted of two biantennary N-glycans in positions A and B (2 x HexNAc₄Hex₅Neu5Ac₂) and one O-glycan in position D (HexNAc₁Hex₁Neu5Ac₂). We explained the absence of this

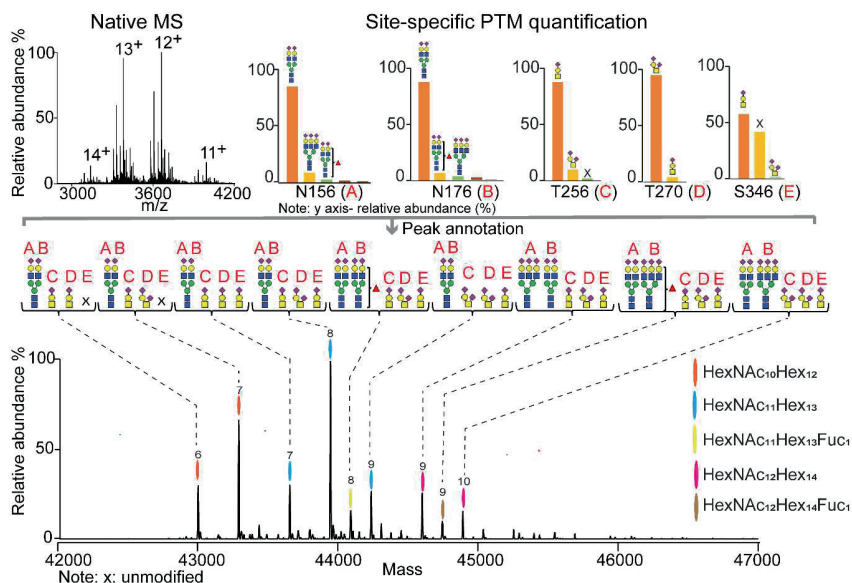


Figure 3. Comprehensive, quantitative and site-specific annotation of fetuin proteoform profiles. The site-specifically annotated zero-charge deconvoluted native mass spectrum of fetuin pooled from human sera of various donors. The overall PTM compositions of the most abundant proteoforms are color coded. Each color represents a glycan composition without the sialic acids; the number of sialic acids attached is marked on the top of each peak. All displayed proteoforms contain one phosphate moiety. The site-specific annotation of the 5 glycosylation sites present on fetuin (A=Asn156, B=Asn176, C=Thr256, D=Thr270, and E=Ser346) was assigned by using our in-house developed software, making use of the integration of the native MS and peptide-centric MS data. The complete list with all annotated proteoforms and their relative abundance can be found supplemental table S2.

proteoform in AHSG*1 above, being due to the complete occupation of position C (Thr256) O-glycan-site. This was in pronounced contrast with position C (Ser256) in AHSG*2, which was found to be occupied only partially. The second major difference between AHSG*1 and AHSG*2 was defined by their different protein backbone mass induced by the two mutations, resulting in the exact mass difference of 15.99 Da (37,177.01 - 37,193.00). This could be extracted already from the two most intensive proteoforms in the native mass spectra of F1

(AHSG*1) and M3 (AHSG*2) displayed in Supplemental Figure S4AB. Although both of these proteoforms had the same total PTM composition, site-specific annotation also revealed a difference in the O-glycan positions occupancy. Donor F1 had O-glycans at positions C and D, whereas M3 at positions D and E. A complete list of site-specifically annotated proteoforms from the three individuals representing different genotypes can be found in Supplemental Tables S3-5. In F1, we could annotate 16 proteoforms (Supplemental Table S3), in M3 15 proteoforms (Supplemental Table S4), and lastly in the heterozygote donor F3 34 proteoforms (Supplemental Table S5). The proteoform profiles differed from each other not only by their complexity, but also by the relative abundances of peaks.

Having annotated the three representative native mass spectra, we attempted a classification of all 10 studied healthy individuals based on their proteoform profiles, following a similar approach as used earlier to classify glycol-engineered erythropoietin variants (47). We initially used unsupervised hierarchical clustering to construct a matrix based on all native mass spectra, as shown in Figure 4A. This already resulted in three distinctive clusters, but also two clear outliers (donors F5 and M5). The first cluster represented fetuins from AHSG*2 donors (orange box). The second (green box) and third cluster (blue box) contain fetuin from AHSG1/2 and AHSG*1 donors, respectively. Upon closer inspection, the native mass spectra of the two outlier samples (F5 and M5), displayed more complex proteoform profiles, with nearly all peaks co-existing in pairs differing from each other by 80 Da (Supplemental Figure S2). From the LC-MS/MS data we could conclude that these two donors have a high occupancy of a second phosphorylation site, namely Ser330 (Supplemental File S2). The presence of this extra phosphate moiety interfered with the genotype classification based on the complete proteoform profiles. Nevertheless, as we above identified characteristic proteoforms defining each genotype, we repeated the classification, but now using only these signature peaks. The result is shown in Figure 4B, where the orange, green and purple boxes represent the genetic variants of fetuin. This data clearly exemplifies that, over a sample size of 10 healthy individuals, distinctive glycosylation patterns can be linked to the distinctive fetuin genotypes.

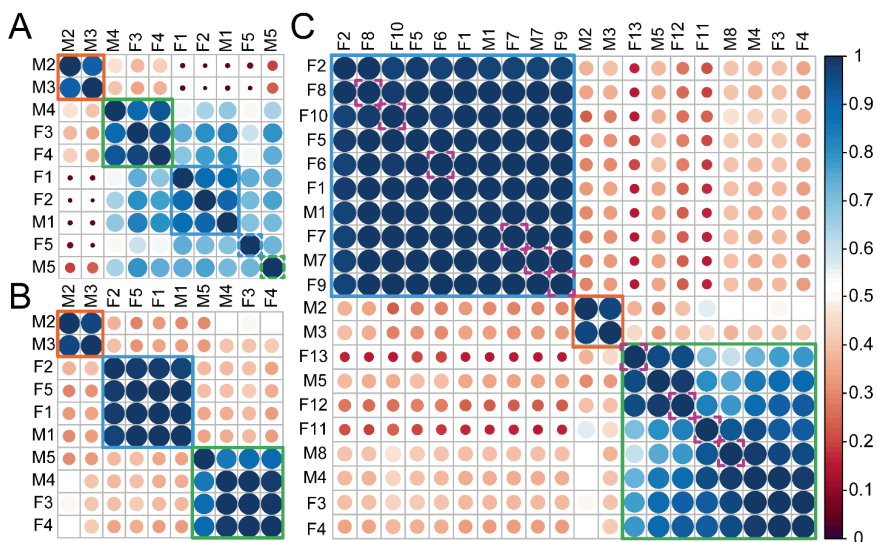


Figure 4. Classification of fetuin proteoform profiles by hierarchical clustering. A. Unsupervised clustering of fetuin proteoform profiles derived from the 10 healthy individuals based on the correlation between their native mass spectra. B. Clustering of fetuin proteoform profiles derived from 10 healthy individuals based on specific signature peaks (see supplemental Figure S4). Color and size of the circles represent the similarity between the proteoform profiles of different genotypes. Orange, blue and green boxes represent fetuin originating from donors representing the AHSG*2, AHSG*1 and AHSG1/2 genotypes, respectively. Comparing the results in A and B the classification in A is distorted by two outliers derived from donor F5 (AHSG*1) and M5 (AHSG*2), respectively caused by a major change in fetuin phosphorylation in these two individuals. C. Clustering of fetuin proteoform profiles derived from 10 healthy and 10 septic patients based on specific signature peaks, where pink dashed line boxes indicate fetuin derived from septic individuals.

Proteoform profiles of fetuin from septic patients donors

Alternation of glycosylation can be a trademark of inflammation, as reported to occur on several acute phase proteins; α -1-acid glycoprotein, haptoglobin, and α -1-antichymotrypsin (34, 35). However, such data for fetuin is sporadic (48). To investigate, whether fetuins' glycosylation pattern also alters upon acute inflammation, we next purified serum fetuin from 10 individual septic patient donors (Table 1) and conducted the same mass spectrometric analysis as before. Our classification strategy quickly classified all 20 native mass spectra (from the 10 healthy and 10 septic patient donors) into the three genotypes (Fig. 4C). This classification could be further corroborated by the peptide-centric LC-MS/MS data (Supplemental Table S6). Next,

we focused on whether we could classify the healthy and septic patient donors. Fetuin obtained from septic patient donors contained some molecular species with a prominently increased intensity originating from proteoforms with fucosylated N-glycans. Zooming in to this, we focused on a specific fetuin fucosylated proteoform and its non-fucosylated variant and used the ratio of their peak intensities to estimate the level of fucosylation (Figure 5A). We calculated from the native mass spectra the average intensities of the two proteoforms using their peaks detected in the 12⁺ and 13⁺ charge states (Supplemental Table S7). The peptide-centric LC-MS/MS analysis allowed us to pinpoint the by sepsis most affected N-glycosylation site as Asn176 (B). The ratios of peak areas of the fucosylated to non-fucosylated peptide containing this site validated the clear trend of fucosylation increase in septic patients (Figure 5B and Supplemental Table S8). This particular change in proteoform profile can thus be used to classify septic patients from healthy donors with good statistical values ($p < 0.006$).

Discussion

Here we combined high-resolution native mass spectrometry with peptide-centric (glyco)proteomics to analyze the medium abundant serum glycoprotein fetuin, which we were able to directly purify from serum of 20 donors, 10 healthy controls and 10 septic patients. From the data, it became apparent that the samples were derived from people with distinct frequently occurring genotypes, that all result in a different set of fetuin proteoforms. We further investigated in detail the variations in proteoform profiles caused by the donors' specific genotypes, providing an interesting example of glycoproteogenomics. We next also extended our analysis to serum fetuins purified from septic patient donors.

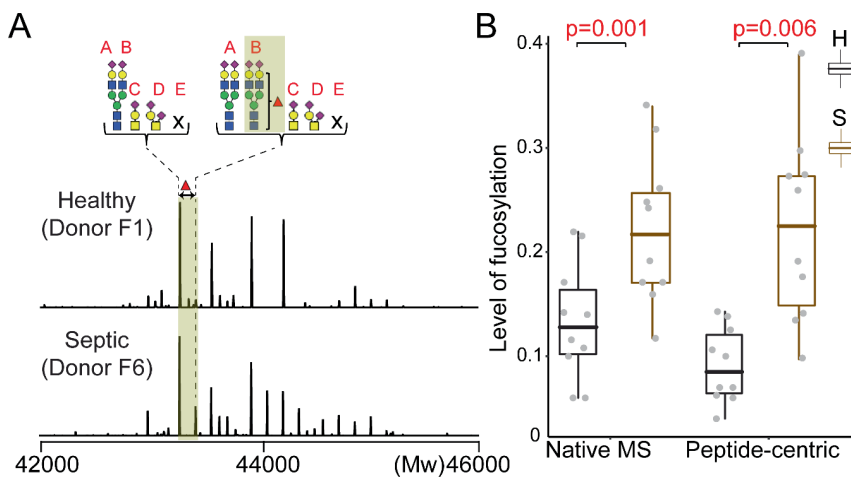


Figure 5. Septic patients display enhanced fucosylation on Asn176 A. Two representative zero-charge deconvoluted native mass spectra of a healthy donor (F1) and septic patient (F6). The average intensities of the site-specifically assigned fucosylated proteoform and its non-fucosylated variant were used for the determination of the relative fucosylation level. The intensities were extracted from the native mass spectra using the corresponding ion signals detected in the 12⁺ and 13⁺ charge states. B. Comparison of the extend of fucosylation obtained from the native mass spectra and peptide-centric analysis. The level of fucosylation in the peptide-centric data was determined by ratios of the peak areas of the fucosylated and non-fucosylated peptide containing the sepsis affected N-glycosylation site Asn176. Both approaches resulted in a statistically significant separation of the healthy and septic patients with p-values of 0.001 and 0.006, respectively. (A=Asn156, B=Asn176, C=Thr256, D=Thr270, and E=Ser346)

*Replacement of Thr in AHSG*1 to Ser in AHSG*2 changes O-glycosylation occupancy*

In both the native MS and LC-MS/MS data we observed that the mutation in position 256, which is a Thr in AHSG*1 and a Ser in AHSG*2, had a dramatic effect on fetuin O-glycosylation, primarily caused by the diminished occupancy of O-glycans on AHSG*2/Ser256. Notably, both a Thr and Ser can be modified by O-glycosylation, but our finding that the O-glycosylation is reduced on fetuin Ser256 compared to Thr256 is in line with a meta-analysis on O-glycosylation occupancy, which reported a considerable preference for Thr over Ser (49). We provide here the first experimental evidence showing how the replacement of Thr for Ser in fetuin changes the occupancy of this site. In general, changes in O-glycosylation may affect protein stability or protection against proteolytic activity (50). Impact of gene mutation,

including SNPs, on protein glycosylation has been reported earlier, albeit only in a few studies (51–53). Notably, such mutations have been associated with altered physiological states, especially the ones leading to disease. Whether different the here reported proteoform profiles of fetuin have biological implications needs to be further investigated.

Coupling between fetuin backbone processing and phosphorylation

The native mass spectra of two healthy donors, F5 and M5, attracted our attention due to their quite distinctive proteoform profiles, representing the two outliers in Figure 4A. We observed that this was primarily caused by the site-occupancy of the additional fetuin phosphorylation site Ser330. Human serum fetuin is predominantly present as singly phosphorylated, with a site-occupancy of nearly 100% at Ser138 (25). Earlier phosphoproteomics analysis has revealed that at least four more Ser residues can be partially phosphorylated (54–56). In a recent study comparing obese *versus* lean individuals, phosphorylation on fetuin Ser330 was attributed with a metabolically active pool of human fetuin, within the larger total pool of circulating fetuin (57). Interestingly, this Ser330 phosphorylation site is located in the 40-amino acid residue long propeptide (also termed “connecting peptide”), which is believed to be lost during posttranslational processing in serum. Data from literature has suggested that fetuin can indeed exist in human serum in two processed forms, with- or without the connecting peptide (58). One putative proteolytic cleavage separating fetuin into two chains occurs at site 340 leading to the removal of an Arg residue. The second proteolytic cleavage releasing the connecting peptide (with the phosphate at Ser330) is Leu300. It is therefore tempting to speculate about a possible cross-talk mechanism between Ser330 phosphorylation and fetuin processing. One proposed functional role of this connecting peptide is in an insulin receptor inhibitory function (INSR) (59, 60). The INSR inhibitory activity of matured human fetuin isolated from plasma (assumed without the connecting peptide) is significantly lower compared to recombinant human fetuin (which is a single-chain protein including the connecting peptide). Since it has been reported that 7-20% of circulating fetuin carries phosphorylation at Ser330 (14), it is

tempting to speculate that a similar portion of human fetuin circulates as a single-chain protein containing the connecting peptide (14, 57). The existence of such a minor, albeit metabolically more active pool of fetuin could be important in the pathology of type 2 diabetes which is connected with acquired or inherited defects in the insulin receptor signaling cascade. Our data suggest a different scenario. We showed in our earlier study and also here, that human serum fetuin exists exclusively in the two-chain form architecture, however it still contains the connecting peptide. The proteolytic cleavage separating fetuin into two chains occurs at aforementioned position 340 and completely removes the Arg residue. Therefore, the modification of phosphorylation site Ser330 does not seem to correlate with the absence/presence of the connecting peptide and a putative regulatory function of the phosphate at Ser330 has likely a different mechanism.

Sepsis specific alterations in the proteoforms of human fetuin

In many diseases, including acute and chronic inflammatory diseases, several acute phase proteins display altered glycosylation (61). Human fetuin has been classified as a negative acute phase glycoprotein, however, whether and how inflammatory processes affect fetuin glycosylation has not been described yet. Here, we observed in fetuin purified from septic patient donors a statistically significant increase of fucosylation on N-glycosylation site Asn176 (B). Notably, two septic patients, M7 and F7, did not show increased fucosylation. Both these patients were considerably younger compared to the other patients (18 and 19 years old). It has been previously hypothesized that N-glycan profiles are affected by age and gender (62–65). Although arguably on a small dataset, we did observe a correlation in our data between the extent of fucosylation and age. However, the number of data points showing the correlation between fucosylation and age was not sufficient to perform a proper statistical evaluation. Nonetheless, this further supports that the discovery and development of biomarkers should seriously take into consideration factors like gender and age of the donors. Whether the

increased fucosylation has a functional consequence or is barely a diagnostic marker of “diseases-state” needs further exploration.

In conclusion, we analyzed here for the first time fetuin samples extracted from twenty individual donors, 10 healthy and 10 suffering from a septic shock. This personalized in-depth proteoform profiling provided a wealth of data on fetuin processing, fetuin phosphorylation and fetuin N- and O-glycosylation. One of the most intriguing findings of our study is that the common fetuin gene polymorphism affect the corresponding proteins’ proteoform profiles.

In most genomics and proteomics studies on fetuin so far consequences of the gene polymorphisms, and fetuin processing have been largely ignored. We feel that apparent discrepancies on fetuin reported in the literature, need to be readdressed, taken into account the donor's individual genotype, physiological state, and fetuin processing state. It has been often suggested that fetuin is a protein sensitive to various physiological states, but whether its clinical potential is strong, deserves to be further investigated. Our data clearly demonstrates how a single gene can lead to a very broad range of proteoforms, which all may be functionally differentially active. At the proteoform level there seem to be no boring genes, as even fetuin, characterized already 75 years ago, still exposes new and exciting features when studied in human individuals exhibiting different genotypes and physiological states.

Acknowledgments

We acknowledge support from the Netherlands Organization for Scientific Research (NWO) funding the large-scale proteomics facility *Proteins@Work* (project 184.032.201) embedded in the Netherlands Proteomics Centre. A.J.R.H. acknowledge further support by the NWO TOP-Punt Grant 718.015.003.

Data availability

The mass spectrometry data have been deposited to the ProteomeXchange Consortium via the PRIDE (66) partner repository with the dataset identifier PXD012465 and 10.6019/PXD012465. As PRIDE does not fully support glycopeptide data Byonic results, in the form of freely available Byonic Preview files, containing all identified and assigned spectra are available at the following link: <https://figshare.com/s/c3ef7dba6c2079e0402f>

References

1. Ix, J. H., Shlipak, M. G., Brandenburg, V. M., Ali, S., Ketteler, M., and Whooley, M. A. (2006) Association between human fetuin-A and the metabolic syndrome: Data from the heart and soul study. *Circulation* 113, 1760–1767
2. Dziegielewska, K. M., Møllgård, K., Reynolds, M. L., and Saunders, N. R. (1987) A fetuin-related glycoprotein (α 2HS) in human embryonic and fetal development. *Cell Tissue Res.* 248, 33–41
3. Denecke, B., Gräber, S., Schäfer, C., Heiss, A., Wöltje, M., and Jahnen-Dechent, W. (2003) Tissue distribution and activity testing suggest a similar but not identical function of fetuin-B and fetuin-A. *Biochem. J.* 376, 135–145
4. Cai, M. M. X., Smith, E. R., and Holt, S. G. (2015) The role of fetuin-A in mineral trafficking and deposition. *Bonekey Rep.* 4, 1–10
5. Jahnen-dechent, L. B. W. (2013) The Role of Fetuin-A in Physiological and Pathological Mineralization. *Calcif. Tissue Int.* 93, 355–364
6. Towler, D. A., Editor, G., Jahnen-dechent, W., Heiss, A., Scha, C., and Ketteler, M. (2011) Fetuin-A Regulation of Calcified Matrix Metabolism. *Circ. Res.* 108, 1494–1509
7. Weikert, C., Stefan, N., Schulze, M. B., Pischon, T., Berger, K., Joost, H. G., Häring, H. U., Boeing, H., and Fritsche, A. (2008) Plasma fetuin-A levels and the risk of myocardial infarction and ischemic stroke. *Circulation* 118, 2555–2562
8. Ketteler, M., Bongartz, P., Westenfeld, R., Wildberger, J. E., Mahnken, A. H., Böhm, R., Metzger, T., Wanner, C., Jahnen-Dechent, W., and Floege, J. (2003) Association of low fetuin-A (AHSG) concentrations in serum with cardiovascular mortality in patients on dialysis: A cross-sectional study. *Lancet* 361, 827–833
9. Wang, H., and Sama, A. E. (2012) Anti-inflammatory role of fetuin-A in injury and infection. *Curr. Mol. Med.* 12, 625–633

10. Yin, L., Cai, W. J., Zhu, L. Y., Li, J., Su, X. H., Wang, X. L., Chang, X. Y., and Sun, K. (2015) Association of plasma Fetuin-A and clinical characteristics in patients with new-onset type 2 diabetes mellitus. *Int. J. Clin. Exp. Med.* 8, 991–999
11. Sun, Q., Cornelis, M. C., Manson, J. A. E., and Hu, F. B. (2013) Plasma levels of fetuin-A and hepatic enzymes and risk of type 2 diabetes in women in the U.S. *Diabetes* 62, 49–55
12. Schäfer, C., Heiss, A., Schwarz, A., Westenfeld, R., Ketteler, M., Floege, J., Müller-esterl, W., Schinke, T., and Jahnen-dechent, W. (2003) The serum protein alpha 2-Heremans-Schmid glycoprotein fetuin-A is a systemically acting inhibitor of ectopic calcification. *J. Clin. Invest.* 112, 357–366
13. Singh, M., Sharma, P. K., Garg, V. K., Mondal, S. C., Singh, A. K., and Kumar, N. (2012) Role of fetuin-A in atherosclerosis associated with diabetic patients. *J. Pharm. Pharmacol.* 64, 1703–1706
14. Kaushik SV, Plaisance EP, Kim T, Huang EY, Mahurin AJ, Grandjean PW, M. S. (2009) Extended-release niacin decreases serum fetuin-A concentrations in individuals with metabolic syndrome. *Diabetes. Metab. Res. Rev.* 25, 427–434
15. Vörös, K., Gráf, L., Prohászka, Z., Gráf, L., Szenthe, P., Kaszás, E., Böröcz, Z., Cseh, K., and Kalabay, L. (2011) Serum fetuin-A in metabolic and inflammatory pathways in patients with myocardial infarction. *Eur. J. Clin. Invest.* 41, 703–709
16. Ismail, N. A., Ragab, S., Abd El Dayem, S. M., Abd ElBaky, A., Salah, N., Hamed, M., Assal, H., and Koura, H. (2012) Fetuin-A levels in obesity: Differences in relation to metabolic syndrome and correlation with clinical and laboratory variables. *Arch. Med. Sci.* 8, 826–833
17. D W Cox, B J Andrews, and D. E. W. (1986) Genetic polymorphism of alpha 2HS-glycoprotein. *Am. J. Hum. Genet.* 38, 699–706

18. Osawa, M., Umetsu, K., Ohki, T., Nagasawa, T., Suzuki, T., and Takeichi, S. (1997) Molecular evidence for human alpha2-HS glycoprotein (AHSG) polymorphism. *Hum. Genet.* 99, 18–21
19. Andersen, G., Burgdorf, K. S., Sparsø, T., and Borch-johnsen, K. (2008) AHSG Tag Single Nucleotide Polymorphisms Associate With Type 2 Diabetes and Dyslipidemia: Studies of Metabolic Traits in 7,683 White Danish Subjects. *Diabetes* 57, 1427–1432
20. Kim, J. G., Kim, H., Ku, S. Y., Kim, S. H., Choi, Y. M., and Moon, S. Y. (2004) Association between human α 2-Heremans Schmidt glycoprotein (AHSG) polymorphism and endometriosis in Korean women. *Fertil. Steril.* 82, 1497–1500
21. Brega A, Scacchi R, Cuccia M, Kirdar B, Peloso G, C. R. (1998) Study of 15 protein polymorphisms in a sample of the Turkish population. *Hum. Biol.* 70, 715–728
22. Cerri, N., and De Ferrari, F. (1991) Genetic polymorphism of Alpha-2-HS glycoprotein in Lombardy (Italy). *Int. J. Legal Med.* 104, 77–79
23. Elzanowski, A., Barker, W. C., Hunt, L. T., and Seibel-Ross, E. (1988) Cystatin domains in alpha-2-HS-glycoprotein and fetuin. *FEBS Lett.* 227, 167–170
24. Araki, T., Yoshioka, Y., and Schmid, K. (1989) The position of the disulfide bonds in human plasma alpha 2 HS-glycoprotein and the repeating double disulfide bonds in the domain structure. *Biochim. Biophys. Acta* 994, 195–199
25. Lin, Y. H., Franc, V., and Heck, A. J. R. (2018) Similar Albeit Not the Same: In-Depth Analysis of Proteoforms of Human Serum, Bovine Serum, and Recombinant Human Fetuin. *J. Proteome Res.* 17, 2861–2869
26. Hoffmann, M., Marx, K., Reichl, U., Wuhrer, M., and Rapp, E. (2016) Site-specific O-Glycosylation Analysis of Human Blood Plasma Proteins. *Mol. Cell. Proteomics* 15, 624–641
27. Spiro, R. G. (2002) Protein glycosylation: nature, distribution, enzymatic formation, and

disease implications of glycopeptide bonds. *Glycobiology* 12, 43R–56R

28. Seo, J., and Lee, K.-J. (2004) Post-translational modifications and their biological functions: proteomic analysis and systematic approaches. *J. Biochem. Mol. Biol.* 37, 35–44
29. Smith, E. R., Ford, M. L., Tomlinson, L. A., Rocks, B. F., Rajkumar, C., and Holt, S. G. (2010) Poor agreement between commercial ELISAs for plasma fetuin-A: An effect of protein glycosylation? *Clin. Chim. Acta* 411, 1367–1370
30. Kirwan, A., Utratna, M., O'Dwyer, M. E., Joshi, L., and Kilcoyne, M. (2015) Glycosylation-Based Serum Biomarkers for Cancer Diagnostics and Prognostics. *Biomed Res. Int.* 2015, 16
31. Gornik, O., and Lauc, G. (2008) Glycosylation of serum proteins in inflammatory diseases. *Dis. Markers* 25, 267–278
32. McCarthy, C., Saldova, R., Wormald, M. R., Rudd, P. M., McElvaney, N. G., and Reeves, E. P. (2014) The role and importance of glycosylation of acute phase proteins with focus on alpha-1 antitrypsin in acute and chronic inflammatory conditions. *J. Proteome Res.* 13, 3131–3143
33. Flood-Nichols, S. K., Kazanjian, A. A., Tinnemore, D., Gafken, P. R., Ogata, Y., Napolitano, P. G., Stallings, J. D., and Ippolito, D. L. (2014) Aberrant glycosylation of plasma proteins in severe preeclampsia promotes monocyte adhesion. *Reprod. Sci.* 21, 204–214
34. Brinkman-Van Der Linden, E. C. M., Van Ommen, E. C. R., and Van Dijk, W. (1996) Glycosylation of α 1-acid glycoprotein in septic shock: Changes in degree of branching and in expression of sialyl Lewisx groups. *Glycoconj. J.* 13, 27–31
35. Brinkman-van der Linden EC, de Haan PF, Havenaar EC, van D. W. (1998) Inflammation-induced expression of sialyl LewisX is not restricted to alpha1-acid

- glycoprotein but also occurs to a lesser extent on alpha-antichymotrypsin and haptoglobin. *Glycoconj. J.* 15, 177–182
36. Yang, Y., Liu, F., Franc, V., Halim, L. A., Schellekens, H., and Heck, A. J. R. (2016) Hybrid mass spectrometry approaches in glycoprotein analysis and their usage in scoring biosimilarity. *Nat. Commun.* 7, 13397
 37. Franc, V., Zhu, J., and Heck, A. J. R. (2018) Comprehensive Proteoform Characterization of Plasma Complement Component C8 $\alpha\beta\gamma$ by Hybrid Mass Spectrometry Approaches. *J. Am. Soc. Mass Spectrom.* 29, 1099–1110
 38. Franc, V., Yang, Y., and Heck, A. J. R. (2017) Proteoform Profile Mapping of the Human Serum Complement Component C9 Revealing Unexpected New Features of N-, O-, and C-Glycosylation. *Anal. Chem.* 89, 3483–3491
 39. Havugimana, P. C., Wong, P., and Emili, A. (2007) Improved proteomic discovery by sample pre-fractionation using dual-column ion-exchange high performance liquid chromatography. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 847, 54–61
 40. Rose, R. J., Damoc, E., Denisov, E., Makarov, A., and Heck, A. J. R. (2012) High-sensitivity Orbitrap mass analysis of intact macromolecular assemblies. *Nat. Methods* 9, 1084–1086
 41. Rosati, S., Rose, R. J., Thompson, N. J., Van Duijn, E., Damoc, E., Denisov, E., Makarov, A., and Heck, A. J. R. (2012) Exploring an orbitrap analyzer for the characterization of intact antibodies by native mass spectrometry. *Angew. Chemie - Int. Ed.* 51, 12992–12996
 42. Bern, M., Caval, T., Kil, Y. J., Tang, W., Becker, C., Carlson, E., Kletter, D., Sen, K. I., Galy, N., Hagemans, D., Franc, V., and Heck, A. J. R. (2018) Parsimonious Charge Deconvolution for Native Mass Spectrometry. *J. Proteome Res.* 17, 1216–1226
 43. Hart, G., Esko, J. D., Kinoshita, T., Cummings, R. D., Schnaar, R. L., Freeze, H. H.,

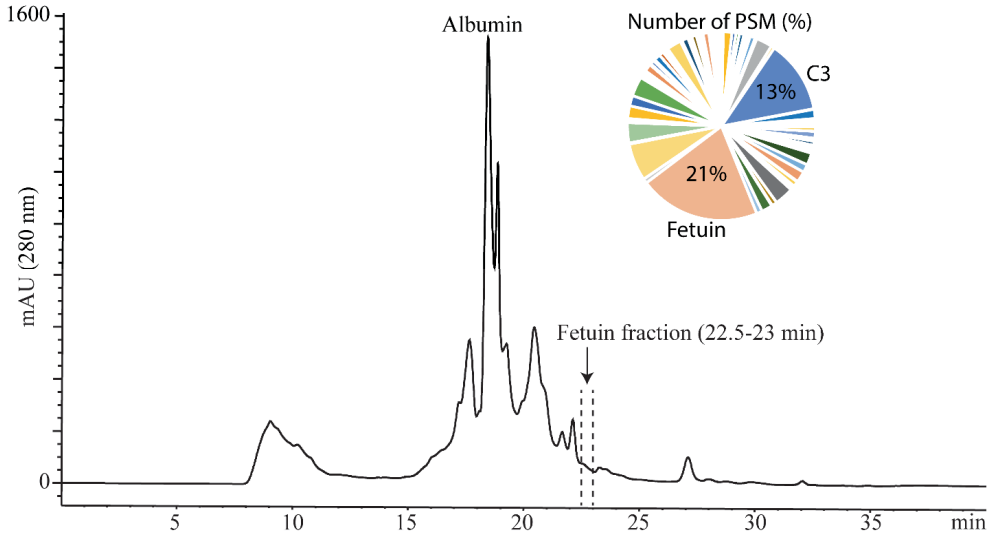
- Etzler, M. E., Bertozzi, C. R., Lütke, T., Dell, A., Vliegthart, J. F., Stanley, P., Prestegard, J. J., Aebi, M., Rudd, P., Packer, N. H., Kanehisa, M., Kornfeld, S., Narimatsu, H., Bolton, E., Varki, A., Darvill, A., Paulson, J., Seeberger, P. H., Perez, S., York, W., Frank, M., Aoki-Kinoshita, K. F., Marth, J. D., Taniguchi, N., and Toukach, P. (2015) Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology* 25, 1323–1324
44. Frese, C. K., Altelaar, A. F. M., Hennrich, M. L., Nolting, D., Zeller, M., Griep-Raming, J., Heck, A. J. R., and Mohammed, S. (2011) Improved Peptide Identification by Targeted Fragmentation Using CID, HCD and ETD on an LTQ-Orbitrap Velos. *J. Proteome Res.* 10, 2377–2388
45. Bern, M., Kil, Y. J., and Becker, C. (2012) Byonic: Advanced Peptide and Protein Identification Software. *Curr. Protoc. Bioinforma.*, Chapter 13:Unit13.20
46. MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., Kern, R., Tabb, D. L., Liebler, D. C., and MacCoss, M. J. (2010) Skyline: An open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 26, 966–968
47. Čaval, T., Tian, W., Yang, Z., Clausen, H., and Heck, A. J. R. (2018) Direct quality control of glycoengineered erythropoietin variants. *Nat. Commun.* 9, 3342
48. Karabudak, A., Giama, N. H., Moser, C. D., Miyoshi, E., and Roberts, L. R. (2017) Identification of fucosylated Fetuin-A as a potential biomarker for cholangiocarcinoma. *PROTEOMICS – Clin. Appl.* 11, 10
49. King, S. L., Joshi, H. J., Schjoldager, K. T., Halim, A., Madsen, T. D., Dziegiel, M. H., Woetmann, A., Vakhrushev, S. Y., and Wandall, H. H. (2017) Characterizing the O - glycosylation landscape of human plasma, platelets, and endothelial cells. *Blood Adv.* 1, 429–442

50. Jentoft, N. (1990) Why are proteins O-glycosylated? *Trends Biochem. Sci.* 15, 291–294
51. Ding, Q., Yang, L., Dinarvand, P., Wang, X., and Rezaie, A. R. (2015) Protein C Thr315Ala variant results in gain of function but manifests as type II deficiency in diagnostic assays. *Blood* 125, 2428–2434
52. Grasbon-Frodl, E., Lorenz, H., Mann, U., Nitsch, R. M., Windl, O., and Kretzschmar, H. A. (2004) Loss of glycosylation associated with the T183A mutation in human prion disease. *Acta Neuropathol.* 108, 476–484
53. Li, S., Iakoucheva, L. M., Mooney, S. D., and Radivojac, P. (2010) Loss of post-translational modification sites in disease. *pacific Symp. Biocomput.*, 337–347
54. Han, G., Ye, M., Zhou, H., Jiang, X., Feng, S., Jiang, X., Tian, R., Wan, D., Zou, H., and Gu, J. (2008) Large-scale phosphoproteome analysis of human liver tissue by enrichment and fractionation of phosphopeptides with strong anion exchange chromatography. *Proteomics* 8, 1346–1361
55. Oppermann, F. S., Gnad, F., Olsen, J. V., Hornberger, R., Greff, Z., Kéri, G., Mann, M., and Daub, H. (2009) Large-scale Proteomics Analysis of the Human Kinome. *Mol. Cell. Proteomics* 8, 1751–1764
56. Zhou, W., Ross, M. M., Tessitore, A., Ornstein, D., VanMeter, A., Liotta, L. A., and Petricoin, E. F. (2009) An initial characterization of the serum phosphoproteome. *J. Proteome Res.* 8, 5523–5531
57. Goustin, A. S., and Abou-Samra, A. B. (2011) The “thrifty” gene encoding Ahsg/Fetuin-A meets the insulin receptor: Insights into the mechanism of insulin resistance. *Cell. Signal.* 23, 980–990
58. Nawratil, P., Lenzen, S., Kellermann, J., Haupt, H., Schinke, T., Mu, W., and Jahnen-dechent, W. (1996) Limited Proteolysis of Human alpha 2 -HS Glycoprotein / Fetuin. *Biochemistry* 271, 31735–31741

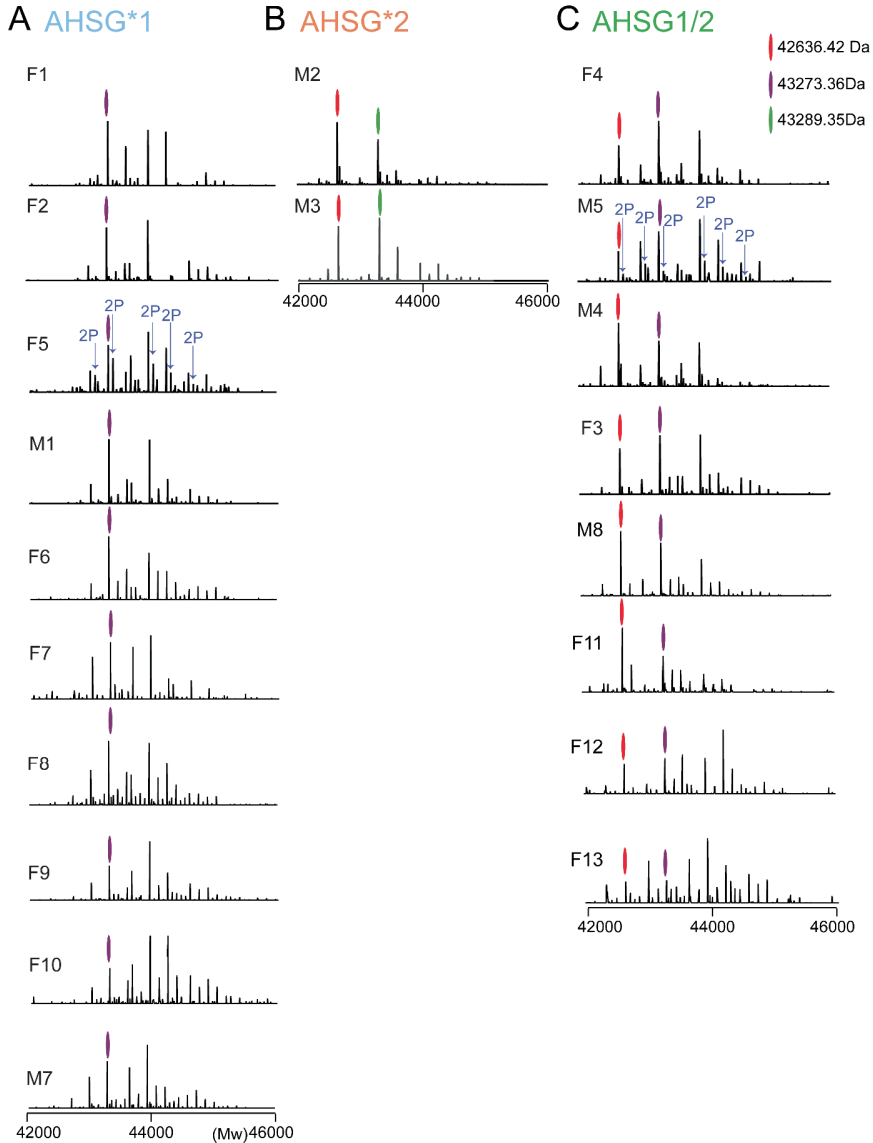
59. Dziegielewska, K., Brown, W., Casey, S., Christie, D., Foreman, R., Hill, R., and Saunders, N. (1990) The Complete cDNA and Amino Acid Sequence of Bovine Fetuin. *J. Biol. Chem.* 265, 4354–4357
60. Srinivas PR, Goustin AS, G. G. (1995) Baculoviral expression of a natural inhibitor of the human insulin receptor tyrosine kinase. *Biochem. Biophys. Res. Commun.* 208, 879–885
61. Mariño, K., Saldova, R., Adamczyk, B., and Rudd, P. M. (2012) Changes in serum N-glycosylation profiles: Functional significance and potential for diagnostics. *Carbohydrate Chemistry* 37, 57-93
62. Ding, N., Sun, H. N., Sun, W., Qu, Y., Liu, X., Yao, Y., Liang, X., Chen, C. C., and Li, Y. (2011) Human serum N-glycan profiles are age and sex dependent. *Age Ageing* 40, 568–575
63. Vanhooren, V., Dewaele, S., Libert, C., Engelborghs, S., De Deyn, P. P., Toussaint, O., Debacq-Chainiaux, F., Poulain, M., Glupczynski, Y., Franceschi, C., Jaspers, K., van der Pluijm, I., Hoeijmakers, J., and Chen, C. C. (2010) Serum N-glycan profile shift during human ageing. *Exp. Gerontol.* 45, 738–743
64. Vanhooren, V., Laroy, W., Libert, C., and Chen, C. (2008) N-Glycan profiling in the study of human aging. *Biogerontology* 9, 351–356
65. Vanhooren, V., Liu, X. E., Franceschi, C., Gao, C. F., Libert, C., Contreras, R., and Chen, C. (2009) N-glycan profiles as tools in diagnosis of hepatocellular carcinoma and prediction of healthy human ageing. *Mech. Ageing Dev.* 130, 92–97
66. Vizcaíno, J. A., Csordas, A., Del-Toro, N., Dianes, J. A., Griss, J., Lavidas, I., Mayer, G., Perez-Riverol, Y., Reisinger, F., Ternent, T., Xu, Q. W., Wang, R., and Hermjakob, H. (2016) 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* 44, D447–D456

Supporting information for Chapter 3

Supplementary Figures

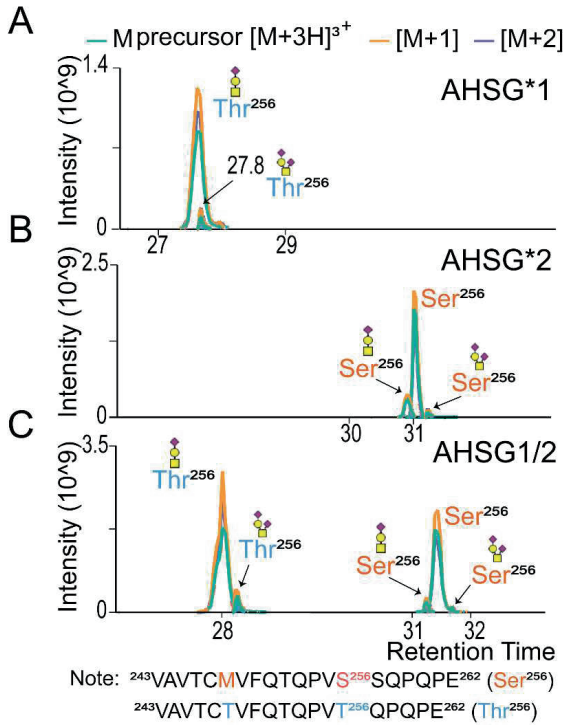


Supplemental Figure S1. Ion exchange based fractionation of human serum from one healthy individual. The fetuin fraction was collected in the time window 22.5-23 min. The proteins in this fractions were identified and quantified by bottom-up proteomics (using number of peptide spectral matches – PSM). The pie chart depicted in the inset reveals that fetuin is not the only, albeit the predominant protein (21%), in this fraction.

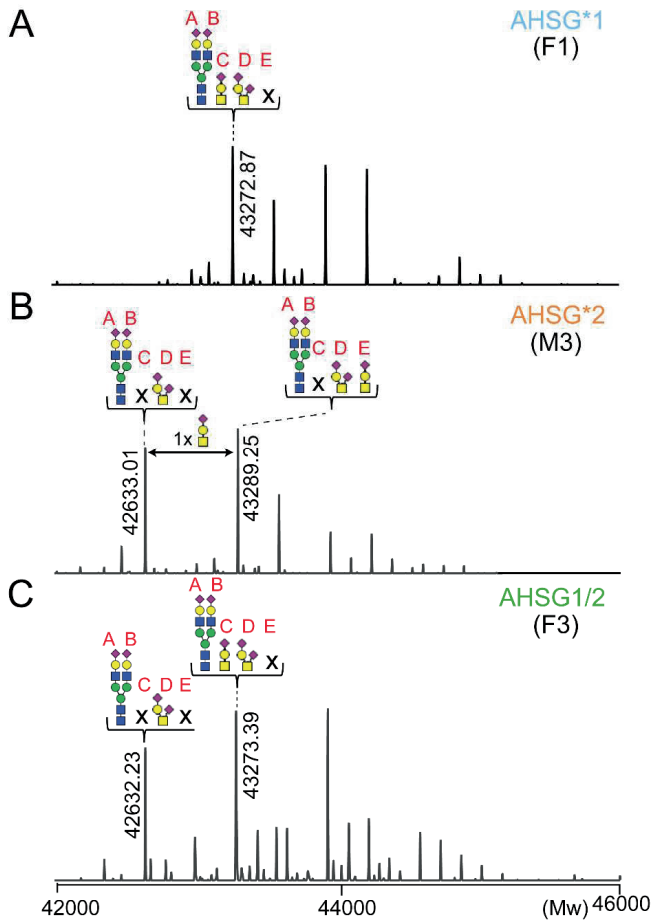


Supplemental Figure S2. Different genotypes lead to different proteoform profiles.

Overview of all zero-charge deconvoluted native mass spectra of human fetuin derived from the 10 healthy and 10 septic patient donors. The spectra are classified into the three fetuin genotypes A. AHSG*1, B. AHSG*2 and C. AHSG1/2 based on the signature peaks assigned with color codes (mass annotated proteoforms). The proteoforms marked with 2P indicate that the PTM composition contains 2 phosphate moieties.



Supplemental Figure S3. Quantification of the peptide signatures containing the mutations and O-glycosylation site Thr/Ser256. Extracted ion chromatograms (XICs) of the peptides containing the O-glycosylation site A. Thr256 on AHSG*1 fetuin, B. Ser256 on AHSG*2 fetuin and C. both Thr256 and Ser256 on AHSG1/2 fetuin.



Supplemental Figure S4. Proteoform profiles originating from individuals representative for the fetuin purified from the serum of A. AHSG*1 donor (F1) B. AHSG*2 donor (M3), and C. AHSG1/2 donor (F3). The depicted glycan combination in a site-specific manner shown in A, B and C corresponds to the signature peaks for the three different genotypes. (A=Asn156, B=Asn176, C=Thr256, D=Thr270, and E=Ser346, X means unmodified glycosylation site.)

The following supplementary tables are not printed because they are oversized.

Supplemental Table S1: Relative quantification of the peptides containing the mutations and O-glycosylation site Thr/Ser256, obtained for all 10 healthy individuals

Supplemental Table S2: List of annotated proteoforms of fetuin pooled from human sera of various donors

Supplemental Table S3: List of annotated proteoforms of fetuin derived from sample F1 (AHSG*1)

Supplemental Table S4: List of annotated proteoforms of fetuin derived from sample M3 (AHSG*2)

Supplemental Table S5: List of annotated proteoforms of fetuin derived from sample F3 (AHSG*1/2)

Supplemental Table S6: Relative quantification of the peptides containing the mutations and O-glycosylation site Thr/Ser256, obtained for all 10 septic individuals

Supplemental Table S7: List of the ratio of the fucosylated proteoform and its non-fucosylated proteoform calculated from the intensities extracted from the native spectra

Supplemental Table S8: List of the ratio of the fucosylated and non-fucosylated peptides containing N-glycosylation site N176 calculated from their peak areas

Supplemental File S1: Human fetuin sample-MS/MS spectra of glycosylation site T/S256

Supplemental File S2: Human fetuin sample F5 and M5- MS/MS spectra of phosphorylation site S330

Supplemental Data

Chapter 4

Glycoproteoform profiles of individual patients' plasma α -1-antichymotrypsin are unique and extensively remodeled in response to a septic episode

Tomislav Čaval^{1,2}; Yu-Hsien Lin^{1,2}; Meri Varkila^{3,4,5}; Karli R. Reiding^{1,2}; Marc J.M. Bonten^{3,4,5}; Olaf L. Cremer^{3,4,5}; Vojtech Franc^{1,2}; Albert J.R. Heck^{1,2}*

¹Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Padualaan 8, 3584 CH Utrecht, The Netherlands

²Netherlands Proteomics Center, Padualaan 8, 3584 CH Utrecht, The Netherlands

³Department of Intensive Care Medicine, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands

⁴Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands

⁵Department of Medical Microbiology, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands

*These authors contributed equally

Authors for Correspondence: Albert Heck, a.j.r.heck@uu.nl (+31302536797)

Manuscript submitted

Abstract

Sepsis and septic shock, the result of a systemic response to an infection, remains the leading cause of death in intensive care units (ICUs). Due to the ambiguous pathogenesis originating from the inflammatory response during sepsis, any contribution to the improved diagnosis and understanding of this process is important. Altered glycosylation patterns of acute-phase proteins have been linked to inflammatory responses. To build further upon such findings, here, we monitored the longitudinal changes in the glycoproteoform profiles of α -1-antichymotrypsin (AACT) extracted from plasma of individual septic patients across four time-points, including post-operative (before sepsis) and post-discharge of the ICU. We used high-resolution native mass spectrometry to qualitatively and quantitatively monitor the multifaceted microheterogeneity of AACT, which allowed us to monitor how changes in the glycoproteoform profiles reflected the patient's physiological state. Although we observed a general trend in the remodeling of the AACT glycoproteoform profiles, each patient exhibited its unique features and responses, providing a resilient proof-of-concept for the importance of personalized longitudinal glycoproteoform profiling.

Introduction

Sepsis is a major cause of morbidity and mortality in hospitals worldwide, especially at their Intensive Care Units (ICUs). It occurs in approximately 2% of all hospitalized patients(1). Mortality rates reach up to 60% for patients experiencing septic shock(1–3). Diagnosis is currently based on symptom scores and microbiological results, which lack both sensitivity and specificity and can lead to significant delays in the initiation of adequate treatment. To date, no single biomarker has been able to diagnose sepsis in an efficient and timely manner. Therefore, further research remains imperative to improve our understanding of the pathophysiology of sepsis and develop new diagnostics and prognostics tools(4).

Of patients admitted to the hospital, blood, plasma, and serum are readily accessible and represent a rich source of proteins used in clinical diagnostics. Consequently, over the last two decades, hundreds of studies focused on mining plasma or serum for biomarkers of all kinds of diseases, notably cancer, diabetes, sepsis, and other inflammatory or infectious diseases(5). Recently, due to the improvements in sample preparations and mass spectrometry-based proteomics technologies, we have been witnessing a renewed interest in plasma proteomics (5, 6). One notable example is a recent large scale study of plasma protein levels in people during longitudinal weight loss, where almost half of all monitored proteins had individual-specific levels that varied significantly between individuals but stayed rather constant in abundance within each individual(7). Focusing on sepsis, a vast body of studies have hinted at close to two hundred proteins associated with the sepsis response (4, 8–13), with the general pathways represented by these proteins, including the acute-phase response (APR), complement activation, and blood coagulation. Unfortunately, many of these proteins are also characteristic biomarkers of several other pathophysiological conditions, and thus not unique for sepsis (4).

During the APR, the concentration levels in plasma of several acute-phase proteins (APPs) change, which is a part of a very complex systemic response to many diseases. Among the most important APPs is C-reactive protein (CRP) that is frequently used in diagnostics as a marker of inflammation and infection(14–16). Measuring CRP values has proven to be very useful in determining treatment effectivity or progress of disease/recovery (17). In addition to changes in protein plasma levels, variations can also occur at a post-translational level, notably in protein glycosylation. These alterations can be specific and are highly dependent on the pathophysiological processes occurring in the cell producing the protein, and on down-stream processing when the protein is secreted into the bloodstream(18–21). As glycosylation is prevalent in proteins found in extracellular environments, the study of glycosylation patterns in proteins from body fluids, such as plasma, has become a valuable source of information to describe cellular processes, including acute inflammation during infection. However, the inherent vast structural complexity of glycoproteins and their proteoform profiles represent a

notorious analytical challenge. Moreover, plasma protein glycosylation is regulated by various factors such as the genetic background of the donor and their environment, making each patient unique(22, 23). We recently reported evidence for this uniqueness, reporting that the proteoform profiles of two abundant plasma APPs, fetuin and haptoglobin, are substantially affected by frequently occurring genetic polymorphisms(24, 25). These observations underline the urgency to investigate plasma proteins in patients individually, i.e., personalized diagnostics also at the protein and proteoform level. Despite many obstacles in identifying specific biomarkers for sepsis, a study from DeCoux et al. has demonstrated an association of specific glycoproteins, such as α -1-antichymotrypsin (AACT), with sepsis survivors(26). We follow up on these important discoveries and here investigate AACT (glyco)proteoforms in unprecedented detail in individual donors that all experienced, and recovered from, a septic episode.

Human AACT is a fairly abundant plasma protein and member of the serpin superfamily of serine proteinase inhibitors that play a key role in the control of several proteolytic cascades. It is mainly produced by the liver and bronchial epithelial cells (27). During an inflammatory response, plasma levels of AACT can be doubled within 16 hours (28), which classifies AACT among the positive-response APPs. Although AACT was described more than half a century ago (29), its precise function remains quite elusive. In general, AACT has a role as an anti-inflammatory agent that inhibits chymotrypsin, cathepsin G, and other proteases (30). ACCT is a heavily glycosylated protein with an apparent molecular mass estimated by SDS-PAGE between 55,000 and 66,000 Da, with the variation attributed to glycosylation microheterogeneity. The protein contains 6 possible N-glycosylation sites and intact glycopeptide analysis confirmed 5 sites as occupied (GlyGen P01011) ; namely asparagine residues at positions 93, 106, 127, 186, and 271 (**Supplementary Figure S1**) (31). It has been shown that N-glycosylation of ACCT is primarily of the complex glycan type; however, a more detailed analysis of the composition of each N-glycan and its heterogeneity has not been provided yet.

Here, we employed state-of-the-art native mass spectrometry (MS) to analyze plasma samples obtained from ten patients, each at four different time-points, who were initially admitted to the ICU for routine observation, yet later developed sepsis. We first monitored the plasma proteome profiles from these patients across the different time points to elucidate changes in protein abundances associated with the sepsis, confirming that several proteins, including C-reactive protein, AACT, and several other APPs, became higher abundant at the onset of sepsis (32–34). Next, we performed quantitative glycoproteome profiling of AACT to monitor potential changes in response to sepsis, using high-resolution native MS (35, 36). Cumulatively, our data revealed that at the onset of sepsis, a substantial glycosylation remodeling occurs on

AACT, whereby it becomes decorated with glycans exhibiting increased branching/LacNAc elongation and fucosylation. Although the glycan remodeling during the septic episode revealed similar features in all patients, each patient carried a remarkable unique glycosylation signature.

Methods

Individual Plasma Sample Collection and Chemicals

From an existing prospective study database, we selected patients who were initially admitted to the ICU for observation following elective surgery, but who developed nosocomial sepsis later during their hospital stay and were consequently re-admitted to the ICU. Leftover plasma was stored at -80°C within 4 hours after collection from the patient for these subjects daily.

Unless otherwise specified, all chemicals and reagents were obtained from Sigma-Aldrich (Steinheim, Germany). Acetonitrile (ACN) was purchased from Biosolve (Valkenswaard, The Netherlands). Sequencing grade trypsin was obtained from Promega (Madison, WI). The Oasis PRiME HLB plate was purchased from Waters (Etten-Leur, the Netherlands).

Plasma sample preparation - Depletion of three abundant proteins

100 µL of each clinical plasma sample (40 in total) was first filtered by using a 0.45 µm filter membrane and subsequently immunodepleted by using a Human plasma Depletion Gravity Column (Good Biotech Corp., Taichung, Taiwan) for the removal of albumin, IgG, and transferrin. The column was first equilibrated with an 8 ml PBS buffer. Next, 100 µL of filtered plasma sample was pipetted on the top of the resin, followed by 1 ml PBS to yield sufficient volume for transferring the plasma through the resin. The flow-through was discarded after 5 minutes. The depleted plasma proteins were eluted and collected from the resin with five column volumes of 1 mL PBS (in total 5 mL), and 5 mL PBS fraction was concentrated down to around 50 µL by using 10 kDa cutoff filter. The concentrated plasma samples were stored at -80 °C until the ion-exchange chromatography separation.

AACT purification from individual plasma samples

An Agilent 1290 Infinity HPLC system (Agilent Technologies, Waldbronn, Germany) consisting of a refrigerated autosampler with a 500-µL injector loop, binary pump, a vacuum degasser, two-column compartment with thermostat, auto collection fraction module, and

multi-wavelength detector, was utilized to purify AACT. The system consisted of a tandem CAT-WAX (PolyWAX LP, 200 × 2.1 mm i.d., 5 μm, 1000 Å; PolyCAT A, 50 × 2.1 mm i.d., 5 μm, 1000 Å) two-stage column set-up. All columns were obtained from PolyLC Inc. (Columbia, USA). The column compartment was kept at 17 °C while the other compartments were cooled to 4 °C. Mobile phase Buffer A consisted of 100 mM ammonium acetate (AMAC)(pH 7.2) in water, and Buffer B consisted of 2.5 M AMAC in water. Typically, 50 μL of depleted plasma sample was mixed with 250 μL Buffer A and injected per run. Elution was achieved using a multi-step gradient, consisting of six transitions with increasing proportions of Buffer B: (step 1; equilibration) 0% B, 0–5 min; (step 2; salt gradient) 0–11% B, 5–6.5 min; (step 3; salt gradient) 11–36% B, 6.5–23 min; (step 4; high salt rinse) 36–100% B, 23–27 min; (step 5; high salt wash) 100% B, 27–31 min; (step 6; restoration) 100–0% B. The flow rate was set to 800 μL/min. The chromatograms were monitored by absorption at 280 nm and time-based fractions collected every 0.5 min from 12-20 min using an automated fraction collector. Human AACT eluted in the fraction collected in between 15.50-16.00 min (**Supplementary Figure 1**).

Native MS Analysis of AACT

For each patient and each time-point, the tandem CAT-WAX fraction containing about 20-40 μg of the AACT protein was first treated with sialidase to remove sialic acids. The desialylated AACT samples were subsequently analyzed on a modified Exactive Plus Orbitrap instrument with extended mass range (EMR) (Thermo Fisher Scientific, Bremen) using a standard m/z range of 1000-10,000 (37). The voltage offsets on the transport multi-poles and ion lenses were manually tuned to achieve optimal transmission of protein ions at elevated m/z . Nitrogen was used in the HCD cell at a gas pressure of $6-8 \times 10^{-10}$ bar. The MS parameters used: spray voltage 1.2-1.3 V, source fragmentation 30 V, source temperature 250 °C, collision energy 30 V, and resolution (at m/z 200) 17,500. The mass spectrometer was calibrated using CsI clusters, as described previously (38).

The masses of the observed glycoproteoforms of AACT were extracted from the zero-charge deconvoluted native mass spectra using intact Mass software (Protein Metrics ver. 3.8) (39). For the analysis of the glycan compositions, the data was processed manually, whereby the glycan structures were retrieved based on known biosynthetic pathways. The average masses used for the glycans were: hexose/mannose/galactose (Hex/Man/Gal, 162.1424 Da), N-acetylhexosamine/N-acetylglucosamine (HexNAc/GlcNAc, 203.1950 Da), deoxyhexose (dHex, 146.1430 Da). Peak intensities in the deconvoluted mass spectra were used for the relative quantification of all the co-occurring AACT glycoproteoforms. Each proteoform was normalized individually in the spectrum so that the sum of all proteoforms amounted to 100%.

Plasma proteome profiling

2 μL of human plasma was taken from each sample of each donor at each time-point and mixed with 50 μL digestion buffer for reduction and alkylation of the proteins. Of note, these plasma proteome analyses were performed without the depletion of any proteins and desialylation. The digestion buffer contained 100 mM Tris-HCl (pH 8.5), 1% w/v sodium deoxycholate (SDC), 5 mM Tris (2-carboxyethyl)phosphine hydrochloride (TCEP) and 30 mM chloroacetamide (CAA). Trypsin was then added for overnight digestion at 37 °C at an enzyme-to-protein ratio of 1:20 (w/w). The next day the SDC was removed via acid precipitation (0.5% trifluoroacetic acid) (TFA). Following acid precipitation, peptides were desalted using an Oasis PRiME HLB plate, then dried and stored at -80 °C until MS analysis.

LC-MS/MS analysis on these peptide samples was performed by coupling an Agilent 1290 Infinity HPLC system (Agilent Technologies, Waldbronn, Germany) to a Q Exactive HF-X mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). The peptides were first trapped with a 100 μm inner diameter 2 cm trap column (in-house packed with ReproSil-Pur C18-AQ, 3 μm) (Dr. Maisch GmbH, Ammerbuch-Entringen, Germany) coupled to a 50 m inner diameter 50 cm analytical column (in-house packed with Poroshell 120 EC-C18, 2.7 μm) (Agilent Technologies, Amstelveen, The Netherlands). The mobile-phase solvent A consisted of 0.1% FA in water, and the mobile-phase solvent B consisted of 0.1% FA in 80% ACN. A 115 min gradient was used as follows: 0-5 min, 100% solvent A; 13-44% solvent B for 95 min; 44-100% solvent B for 5 min; 100% solvent B for 5 min; 100% solvent A for 10 min. Peptides were ionized using a spray voltage of 1.9 kV and a heated capillary. The mass spectrometer was set to acquire full-scan MS spectra (375-1600 m/z) for a maximum injection time of 20 ms at a mass resolution of 60,000 and an automated gain control (AGC) target value of 3e6. Up to 15 of the most intense precursor ions were selected for tandem mass spectrometry (MS/MS). HCD MS/MS (200-2000 m/z) acquisition was performed in the HCD cell, with the readout in the Orbitrap mass analyzer at a resolution of 30,000 (isolation window of 1.4 Th) and an AGC target value of 1e5 or a maximum injection time of 50 ms with a normalized collision energy of 27%.

MaxQuant (software version 1.6.3.4) (40) was used to analyze the LC-MS/MS raw files. MS/MS spectra were searched against the Swiss-Prot FASTA database (release date: Feb 2018, 20 412 entries, taxonomy: *Homo sapiens*) with the following modifications: fixed cysteine carbamidomethylation, variable methionine oxidation. Enzyme specificity was set as C-terminal to arginine and lysine with a maximum of 2 missed cleavages. The searches were performed using a precursor mass tolerance of 8 ppm and a fragment mass tolerance of 20 ppm followed by a 1% false discovery rate (FDR) for protein and peptide levels with a minimum

length of six amino acids for peptides. For label-free quantification, the "match between run algorithm" in the MaxQuant quantification suite was used. Label-free protein quantification (LFQ) was performed for unique peptides with a minimum ratio count > 2.

Results

All plasma samples were obtained via intensive care units (ICU) with patients belonging to the MARS cohort. The MARS cohort includes patients above the age of 18, who have been admitted to the ICUs of the AMC Amsterdam, NL (study period 2011-2013) or the UMC Utrecht, NL (study period 2011-present day) except for elective cardiothoracic surgical patients with an expected length of stay shorter than 24 hours. For the current study cohort, we focused on patients who had been admitted to the ICU on multiple occasions: a first occasion was a "routine" observation after elective surgery with generally an uncomplicated stay shorter than two days; the second and subsequent occasions were ICU-readmission after being suspected/diagnosed with sepsis or a septic shock.

We selected ten individual septic patients whose plasma samples had been collected across four time-points (T1-T4). A summarizing overview of the patient's plasma collection details is depicted in **Figure 1**. The samples at T1 were obtained from the patients who had had elective surgery. Samples T2 were obtained during the suspected septic episode-based, as evidenced by a substantial rise in the plasma abundance of the C-reactive protein, whereby plasma samples for T3 were collected the next day (i.e., after T2). T4 was obtained following the discharge from the ICUs. Notably, there is a difference in the timing of plasma collection for each patient, as it was based on pathophysiological state, not on the exact number of days (see Figure 1). Some patients were already released from the ICU after two days, while some stayed for a couple of weeks.

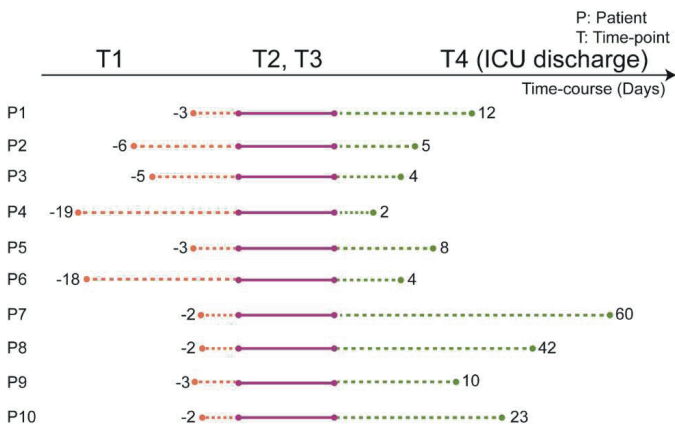


Figure 1. Timing of the plasma sample collection based on the patient's physiological state. Plasma samples were collected from ten individuals each at four subsequent time-points. T1 represents the first plasma sample collection point taken for each patient undergoing elective surgery. The time-points T2 and T3 represent plasma collections at two consecutive days following the assumed septic episode. The plasma from T4 was collected when the patients were discharged from the ICU. Corresponding numbers represent the elapsed number of days between the collection of the samples.

Individual plasma proteome profiles confirm known trends associated with sepsis, including AACT

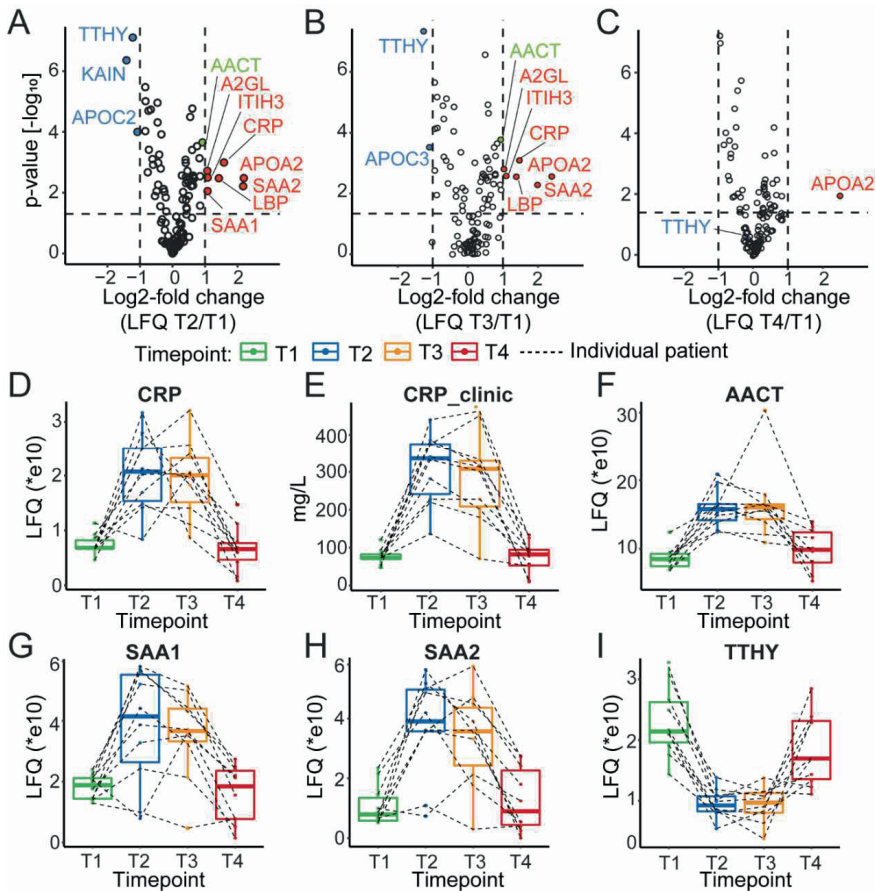
Considering that sepsis is a multifactorial complex syndrome involving a multitude of cell types and organs, it comes to no surprise that clinical definitions of sepsis are based on the patient's symptoms instead of the underlying molecular signatures(41, 42). Several plasma and serum proteomics studies on sepsis have corroborated that the fundamental molecular mechanisms are indeed complex and diverse (12, 43–45).

We first monitored the plasma proteome profiles of each septic patient, focusing just on the most abundant proteins. Using label-free quantitative (LFQ) proteomic profiling, we obtained abundance profiles of about the 200 most abundant plasma proteins (**Supplementary Table 1**). We used the LFQ values of the proteins in T1 to normalize a baseline for subsequent monitoring of the quantitative changes in the plasma proteome (**Supplementary Table 2**). Our analysis revealed about ten significantly higher abundant proteins at T2 and T3 when compared to T1 (**Figure 2A-B**). These upregulated proteins included several acute-phase response proteins (APPs) (ITIH3 - inter- α -trypsin inhibitor, CRP – C-reactive protein, APOA2 - apolipoprotein A2, SAA1 - serum amyloid A1, SAA2 - serum amyloid A2, A2GL - leucine-rich α -2-glycoprotein, and LBP - lipopolysaccharide-binding protein). Each of these displays a wide range of different functionalities, such as serine protease inhibitor (ITIH3), lipid transport (APOA2), and cholesterol transport (SAA1, SAA2). Both the upregulated and downregulated proteins observed in our analysis are commonly associated with sepsis, as well as other inflammatory physiologies, validating that all our patients indeed show the hallmarks of a septic episode(34, 46–49). Finally, following the discharge from the ICU patient plasma proteome profiles seemingly returned close to baseline for most of these proteins (**Figure 2C**).

Since the clinical data on the patient of the MARS cohort contained information on the abundance level of the CRP in plasma (**Supplementary Table 3**), we could compare the trends in CRP level per individual patient when measured with proteomics and ELISA (**Figure 2D and 2E**). Additional examples of LFQ trends of positive (SAA1 and SAA2) as well as negative (TTHY) acute phase proteins are depicted in **Figure 2G-I**. The abundance trend of α -1-antichymotrypsin (AACT), which we selected later for in-depth glycoproteoform profiling due

to its suggested active role in sepsis survivors (26), followed roughly a similar trend as CRP (**Figure 2F**), with an increase in abundance during sepsis in T2 and T3 and reverting close to baseline in T4 for all monitored patients.

As mentioned above, the plasma levels of CRP are often used to assess the condition of septic patients as one of the markers for potential release from the ICU. Our data show that such assays could be complemented by assays monitoring the levels of other APPs, notably ITIH3, CRP, APOA2, SAA1, SAA2, A2GL, LBP, but also AACT.



across all four time-points. The dashed lines connect for each individual patient its corresponding time-points.

Remodeling of patients α -1-antichymotrypsin glycoproteoform profiles during sepsis

Confirming that AACT protein levels follow a similar trend, with a notable increase in abundance upon sepsis, reverting to baseline at T4, we next questioned whether there would also be changes in the glycoproteoform profiles of patients undergoing sepsis and whether such trends would be shared or unique for all studied patients. We chose to focus on AACT as it represents a plasma glycoprotein harboring some of the most elaborate glycosylation profiles, being modified by several biantennary type N-glycans (50).

We and others have shown that high-resolution native mass spectrometry profiles can provide valuable information about protein glycosylation and other post-translational modifications (PTMs), but also on other structural variations, such as genetic variants or mutations (25, 35, 51–53). We thus used such a native MS approach to investigate the structural heterogeneity of AACT in the context of the above-described study design. Although AACT is one of the more abundant proteins in plasma (0.4 mg/L)(54), the high dynamic range of plasma protein levels hampers the facile isolation of AACT by IEX directly from plasma. We overcame this problem by first depleting the three most abundant plasma proteins (albumin, IgG, and serotransferrin) using a dedicated column and semi-automated set-up for immunodepletion, starting with around 100 μ L of a plasma sample. After the initial depletion, we subjected the samples to IEX and collected the fraction containing AACT (**Supplementary Figure S2**). This fraction was subsequently analyzed by native MS. Similar to the observations made by Wu et al. on acid glycoprotein (35) and Tamara et al. (25) on haptoglobin, the highly heterogeneous and complex AACT glycosylation profile precluded us from making a full spectral annotation. The high structural heterogeneity is primarily caused by the high frequency and abundance of sialic acid moieties. They do not only expand the proteoform profiles immensely but also hamper the annotation, as mass shifts caused by two fucosylation moieties or one sialylation moiety are hard to disentangle due to their similar masses (i.e., 291.1 and 292.1 Da, respectively). Hence, in line with the previous studies, we decided to treat the AACT samples with sialidase to remove all sialic acid moieties, clarifying the resulting native mass spectra (**Supplementary Figure S3A-B**) and proved to be very reproducible (**Supplementary Figure S3C**). The deconvoluted native mass spectra of desialylated AACT from T1 of one donor contained around 30 distinct mass signals corresponding to various glycoproteoform variants (**Supplementary Table 4**). Most ion signals center around 55 kDa, whereas the "bare" AACT backbone mass is 45265.82 Da, leaving around 10 kDa of the measured mass corresponding to glycans. Notably, the ion signals of the non-desialylated AACT indicate a Mw of AACT of around 58 kDa, indicating

that we release on average 13-15 sialic acids using Sialidase treatment. The most abundant signal in the desialyated AACT spectrum corresponds to the AACT protein modified with five N-glycans. A possible composition of these five glycans to come to the measured mass is depicted in **Figure 3A**. Next, if we compare T1 with T2 and T3, we see a substantial shift in distribution towards higher masses, caused by increased fucosylation and additional branching of the N-glycans (**Supplementary Figure S4 and Supplementary Table 5**). Surprisingly, this glycan remodeling continues even further onto T4, where we reproducibly observed even greater extent of glycan remodeling (**Figure 3B-F and Supplementary Figure S5**). Since this extensive glycan remodeling, with the enhanced occurrence of fucosylation and additional branching increases the glycoproteoform mass, we decided to determine the weighted average mass (WAM) to effectively and quantitatively visualize these substantial mass shifts. This average mass is a single value representing the WAM of all annotated AACT glycoproteoforms (see Methods for parameters and native spectra annotation). Using this concept, it becomes immediately apparent that across all patients, we observe a substantial initial glycan remodeling in T2 and T3 when compared to T1 (**Figure 3C and 3D**). Additionally, in contrast to the observation that the AACT protein abundance returned to baseline levels comparable to T1 at T4 (**Figure 2F**), we observed a further increase in glycan remodeling at T4 when compared to all earlier three time points (**Figure 3C, 3D and Supplementary Table 5**). From this data it appears that an acute inflammatory event, in this case sepsis, induces a longer-term glycosylation remodeling on AACT, even though that the abundance levels of AACT (and CRP) return to baseline levels and the patients have been released from the ICU.

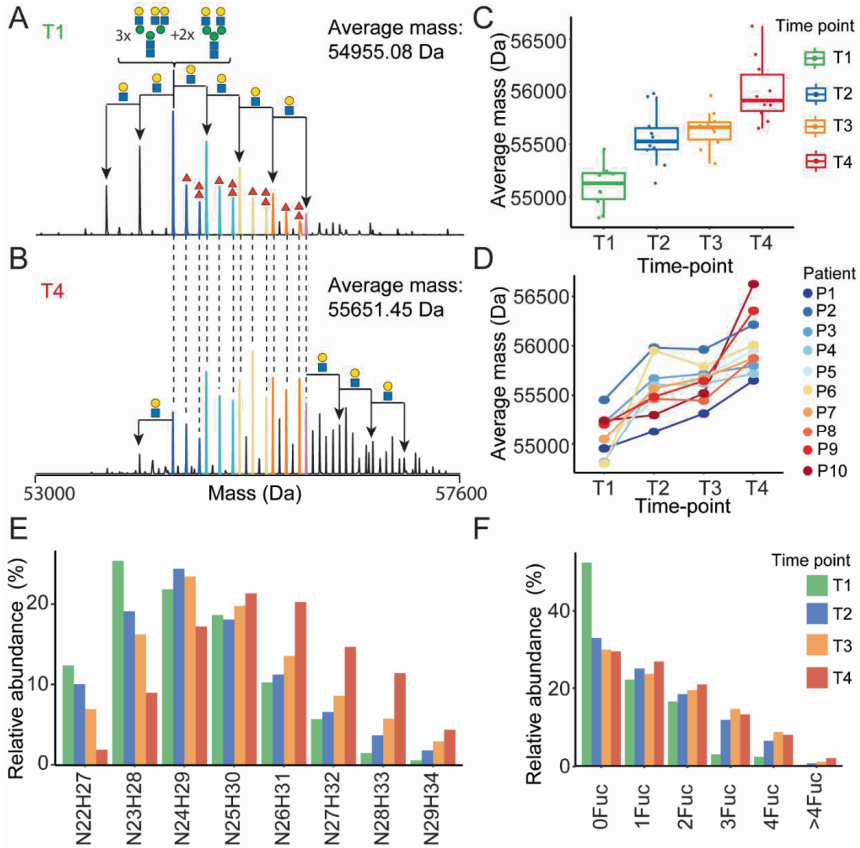


Figure 3. Remodeling of α -1-antichymotrypsin glycosylation following sepsis. Illustrative deconvoluted native mass spectra of AACT in T1 and T4 are shown in A) and B), respectively (for patient 1). The likely glycan composition of the most abundant signals are annotated in A), whereby the GlcNAcGal shifts from the most abundant peak are denoted with black arrows. Red triangles denote the presence of additional fucose residues. Overlapping peaks in the spectra between A) and B) are color-coded, and the weighted average mass (WAM) is denoted in the upper right corner. WAMs for AACT for all patients and time-points are shown in C) represented in Boxplots, and concomitant "spaghetti" plots connecting each patient's data are depicted in D). Relative abundances of AACT glycoproteoforms of patient P1 with the same HexNAcHex (N=HexNAc, H=Hex) or Fuc count are depicted in E) and F), respectively. An overview of the abundances of these glycoproteoforms in AACT for all other patients is provided in **Supplementary Figure 2**.

Despite an evident trend in the increased WAM of AACT across four-time points observed in all patients, we observed significant differences in glycoproteoform profiles between the different patients. This variety means that the T1 baseline for each patient is already quite different. This is also visualized in **Figure 3D**, wherein we plotted the trends for each patient separately. For instance, although both P1 and P2 show a similar continuous trend in the mass

increase over time, the end-point in mass for P1 (i.e., at T4) is similar to the mass for P2 at the start (i.e., at T1). Such trends would, therefore, be missed if we pooled all ten samples, making our study a case for personalized longitudinal glycoproteoform profiling.

Personalized glycoproteoform profiling to assess inter-individual glycosylation heterogeneity

To further assess inter-individual glycosylation heterogeneity in AACT, we utilized our previously introduced concept of "spectral barcoding" (51) wherein the deconvoluted mass spectra of AACT from each patient and T1-T4 time-points are used as glycoproteoform signatures (i.e., barcode). These signatures are subsequently directly compared with each other for mass spectral similarity using the Pearson correlation coefficient. As entries, we have the 40 mass spectra (from 10 patients at four-time points) and calculated the correlation coefficient between all these spectra. In **Figure 4**, an overview of all this data is given in the form of a correlation matrix, whereby spectra that are alike (having a high Pearson correlation) cluster together. From the upper left quadrant of this matrix, it is apparent that the spectra of all patients at T1 cluster well together, with notable exceptions (e.g., P2). However, except for some samples taken at T4, i.e., from patients P10, P2, and P9, which show a distinctive cluster, we did not observe a clear separation at any of the other time-points. This representation corroborates the observation that each patient has a unique baseline in their AACT glycoproteoform profiles, and the onset of sepsis causes extensive remodeling of the AACT glycoproteoform profile. Intrigued by this observation of continuous glycosylation remodeling even up to T4 (when patients are released from the ICU, and AACT and CRP levels have returned already to "normal"), we focused next solely on the differences between the T1 and T4 samples.

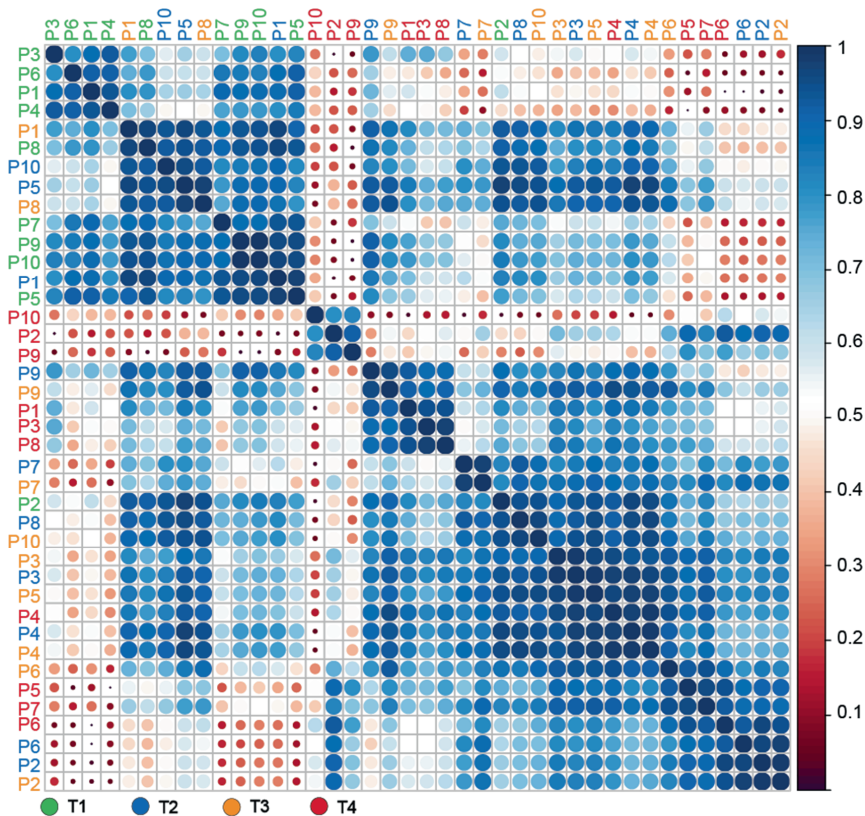


Figure 4. Correlation matrix of α -1-antichymotrypsin glycoproteoform profiles for all patients at all time points. Native mass spectra for each patient and time-points (10×4 spectra) were correlated with each other and clustered based on the extracted Pearson correlation. Patients are color-coded according to the legend below the matrix. The extracted correlation is indicated both by the scale bar on the right, with dark blue representing the highest correlation, and the size of the circles where bigger circles represent higher correlation.

Constructing a similar correlation matrix by taking only the data of the patients at T1 and T4 into account, we observed a clear separation between those two-time points (**Figure 5**). We again observed one notable exception, T1 of patient 2, which seemingly falls into the late-time point cluster. Notably, patient 2 already showed the most significant WAM at T1, as depicted in **Figure 3C-D**. Additionally, we noticed that the spectra of all patients taken at T4 were separated into two partially overlapping clusters, indicated with a purple and brown box in **Figure 5**, respectively. We first assessed whether this separation might be due to the differences in the sample collection points of T4 (see **Figure 1**) (e.g., T4 for patient 4 was collected only four days following T3, while for patient 7 this time difference was 60 days), but this did turn out not to be a determining factor. The most evident factor correlating to the distinction between

the data in the purple and brown box seemed to be the extent of observed glycan remodeling at T4, as represented by the increase in WAM values. The patients in the brown cluster display a lower increase in WAM when compared to those in the purple cluster. In agreement, the T4 samples of the patients in the brown box also correlate somewhat better with cluster 1, representing the data of nearly all T1 samples.

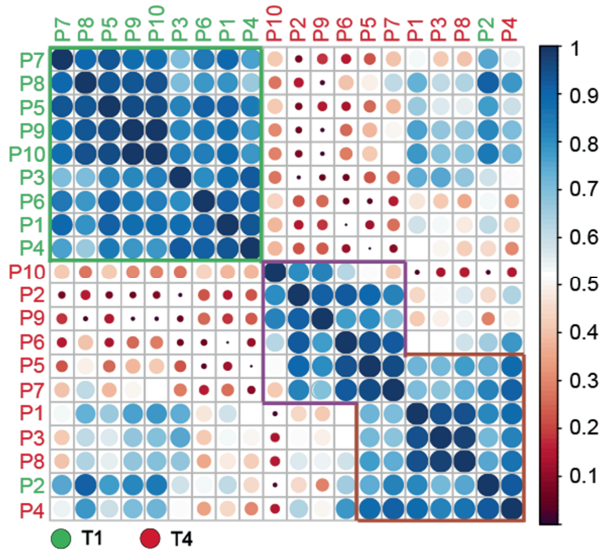


Figure 5. Correlation matrix of α -1-antichymotrypsin glycoproteoform profiles for all patients at time points T1 and T4. Native mass spectra for each patient and time-points were correlated with each other and clustered based on the extracted Pearson correlation. Patients are color-coded according to the time-point with green corresponding to T1 and red to T4. Highlighted are three distinctive clusters coming out of this analysis: green, purple, and brown. The most likely feature distinguishing these clusters is the average weighted masses, which increase for the data from the green box, *via* the brown to the purple box, respectively.

To further illustrate the inter-individual diversity in AACT glycoproteome profiles, the deconvoluted native mass spectra of selected samples (T1 and T4 pairs) are shown in **Figure 6**. For instance, patient 2 exhibits already a relatively complex and high mass glycoproteome profile at T1, similar to the profiles observed at T4 for some of the other patients (**Figure 6**). Also, the extent of glycan remodeling can vary extensively, whereby especially P10 shows a very high extent of remodeling (**Figure 6**). This data further calls for more personalized approaches in (glyco)proteome profiling.

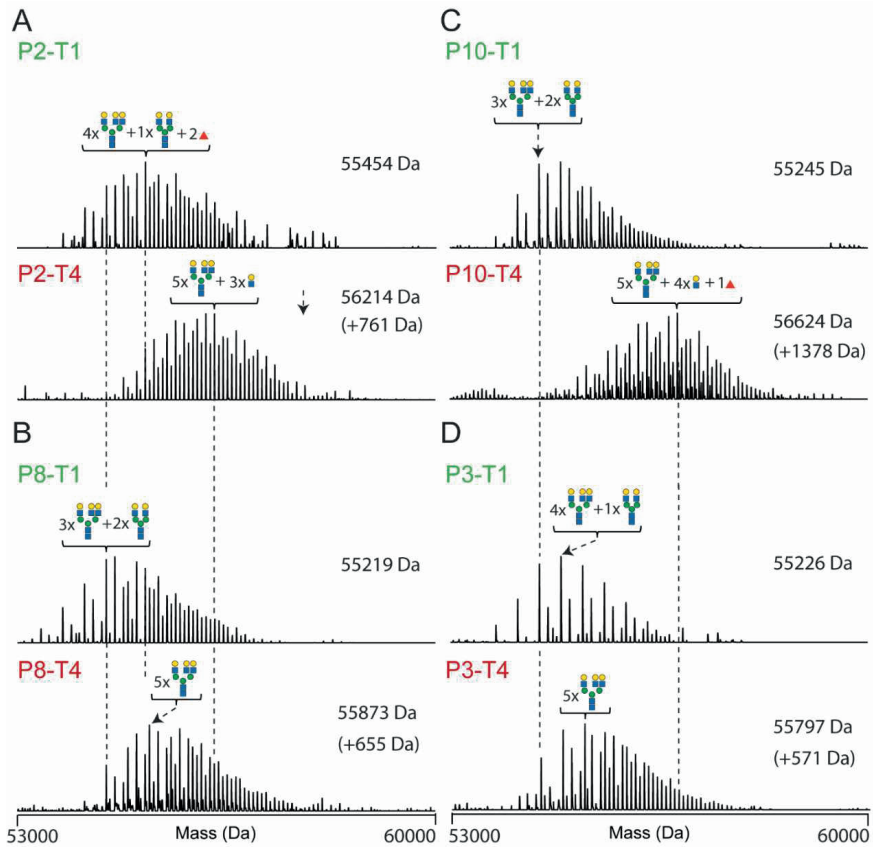


Figure 6. Examples of deconvoluted native mass spectra of AACT illustrating the extensive glycoproteoform remodeling observed between T1 and T4 and unique differences in between patients. Deconvoluted native mass spectra are shown for four patients, each at two-time points (t1 and T4). Each time on top of the panel is the spectrum for T1 while at the bottom, the one of T4. Spectra of patient P2 are shown in A, P8 in B), P10 in C), and P3 in D). The most abundant peak in each spectrum is annotated with one of the most likely glycan compositions. Dashed lines connect peaks of the same mass as a visual guide.

Discussion

In this study, we investigated by proteomics approaches small amounts of plasma obtained from ten patients longitudinally at four time-points throughout a non-lethal septic episode. We first confirmed that this acute inflammatory event resulted in the temporally higher plasma abundance of several positive acute-phase proteins (APPs) that returned to normal levels following partial recovery, whereby patients were released from the ICU. These proteins included the often-used clinical marker CRP, but also, for instance, SAA1, SAA2, and AACT.

Next, we obtained by native mass spectrometry the glycoproteoform profiles of one of these proteins, AACT. We focused on AACT as previous study demonstrated upregulation of AACT glycopeptides is sepsis survivors(28). In our study, we observed a substantial remodeling in its glycosylation already at the onset of the septic episode. Remarkably, this remodeling prolonged even much further, extending beyond the time-point of release of the patients from the ICU. These changes in AACT glycosylation originate primarily from continuous increases in *N*-glycan branching and additional fucosylation. The observed prolonged remodeling of the glycosylation of AACT is thus in sharp contrast to the trend observed for the plasma abundance of AACT, which reverted to “normal” at T4, i.e., when patients were released from the ICUs (Figure 7).

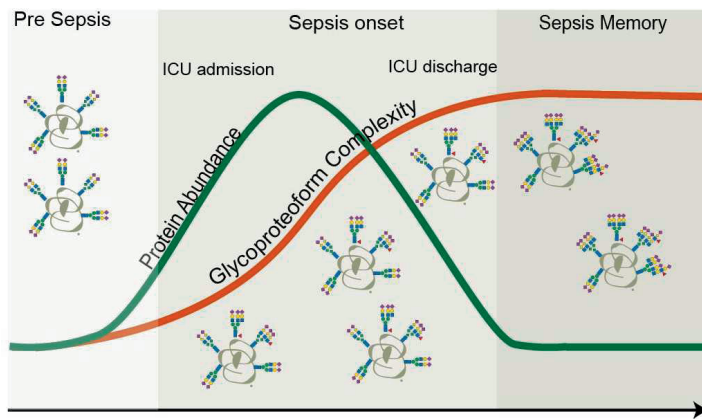


Figure 7. Protein versus proteoform profiling. In this study, both the protein abundance as well as proteoform profile of the plasma protein AACT were monitored longitudinally in ten patients that experienced a septic episode. As AACT is a positive acute-phase protein, its concentration increased during the septic episode and returned to baseline, following the same trend in time as the clinically used marker CRP. Upon return to baseline of CRP (and thus also AACT), patients are typically released from the intensive care unit. At the start of the septic episode also the glycoproteoform profile of AACT started to change and become remodeled and more complicated by the accumulation of additional branches, LacNAc elongations, and additional fucose moieties. Glycosylation remodeling proceeds even after the protein levels have gone back to “normal,” and patients are released from the ICU. This finding means that a memory effect of the septic episode is still present in the patients’ AACT days to weeks after the release from the ICU.

It is well-established that an inflammatory event, regulated by cytokines (predominantly IL-6), can cause differential glycosylation processing during glycoprotein synthesis in the liver(57). It is important to note that glycan remodeling can occur independently of protein synthesis, functionally decoupling protein from glycoproteoform abundance levels as we indeed see on plasma AACT. Our findings here on AACT also are in line with early observations reported on

the α -1-acid glycoprotein. These studies showed that acute inflammation induces action of inflammatory cytokines affecting the expression of the α 1-3 fucosyltransferase enzyme in the liver, resulting in increased levels of sialyl-Lewis X on the α -1-acid glycoprotein(58–60).

One possible reason for the here-observed extensive glycosylation remodeling might be a defense mechanism to protect the patient against immune over-reaction. For instance, aberrant leukocyte activation and recruitment in host tissues can result in the breakdown of the vascular endothelium, organ failure, and death(61, 62). Since sialyl-Lewis X and E-selectin mediate leukocyte extravasation, expression of the same epitope on APPs might represent a way to partially inhibit such interactions(58).

Concerning the observed increase in AACT fucosylation, we hypothesize that these fucose moieties are situated in the antennae of the glycans rather than being attached to the chitobiose core, based on the facts that APPs are seldom core-fucosylated (52, 63). This location of the fucose moieties is in line with the observed concomitant increase in *N*-glycan branching, which results in many more acceptor sites for antennary fucosylation, thus increasing the avidity of sialyl-Lewis X epitope presentation. Although the functional role of *N*-glycosylation in AACT is still elusive, there are a few studies reported on the somewhat similar plasma serine protease, alpha-1-antitrypsin (AAT) (64, 65). *N*-Glycans decorating ATT have a multifaceted function and play a role in protecting the protein from proteolysis and also preventing aggregation(66, 67). Even if this would also be true for AACT, it is unlikely that the glycosylation changes of AACT during sepsis progression observed here influence these potential functions of AACT glycans in particular. Nevertheless, the significant changes in glycoproteoform profiles of AACT could have an impact on its interaction properties with substrates and other proteins. This effect would not be unprecedented as Wu et al. and Tamara et al. described that *N*-glycan heterogeneity might regulate interactions between plasma proteins, notably, between haptoglobin and hemoglobin(27, 37).

In conclusion, we demonstrate that monitoring, at high-resolution, proteoform profiles of individual glycoproteins, such as AACT, provides valuable personalized information complementing traditional measurements of plasma concentrations of APPs for monitoring sepsis progression. Focusing on just this single protein, we reveal that each person's AACT glycoproteoform profiles are unique. We expect this to be true for many other plasma glycoproteins, as corroborated by our recent study on plasma fetuin, showing that its glycoproteoform profiles are affected not only by physiological states but also by frequent genetic polymorphism (26). APPs represent a group of proteins in which concentration levels are highly responsive to inflammatory events during sepsis and other diseases. Our findings on

AACT suggest that next to the rate of synthesis of these proteins, we should be critical of their structural changes and dynamics since these are unique in patients and may provide a deeper understanding of complex pathologies such as sepsis.

Data Availability

Proteomics data have been deposited to MASSIVE repository MSV000086041.

Access for reviewers: MSV000086041_reviewer

Password: hecklab20200831

All native MS data annotations are contained within this manuscript and Supplementary File 1.

Author Contribution

T.Č, Y-H.L, V.F, and A.J.R.H designed research and planned experiments. T.Č, Y-H.L, V.F performed all experiments and data analysis. T.Č, V.F, and A.J.R.H wrote the original draft, which was read and approved by all co-authors. K.R.R contributed to study design. M.V., M.J.M.B, and O.L.C. provided the MARS sepsis cohort samples and selected the patient population used in this study. All authors read, edited, and approved the final manuscript.

Acknowledgments

We acknowledge support from the Dutch Research Council (NWO) funding the Netherlands Proteomics Centre through the X-omics Road Map program (project 184.034.019) and the EU Horizon 2020 program INFRAIA project Epic-XS (Project 823839). Additional support for this work came through the TOP-Punt Grant 718.015.003, and the TA-project grant 741.018.201, and K.R.R. acknowledges support from NWO Veni project VI.Veni.192.058.

References

1. Prescott, H. C., and Angus, D. C. (2018) Enhancing recovery from sepsis: A review. *JAMA - J. Am. Med. Assoc.* 319, 62–75
2. Álvaro-Meca, A., Jiménez-Sousa, M. A., Micheloud, D., Sánchez-Lopez, A., Heredia-Rodríguez, M., Tamayo, E., and Resino, S. (2018) Epidemiological trends of sepsis in the twenty-first century (2000-2013): An analysis of incidence, mortality, and associated costs in Spain. *Popul. Health Metr.* 16, 4
3. Chang, H. J., Lynn, C., and Glass, R. M. (2010) Sepsis. *JAMA - J. Am. Med. Assoc.* 304, 1856
4. Martin, G. S. (2012) Sepsis, severe sepsis and septic shock: Changes in incidence, pathogens and outcomes. *Expert Rev. Anti. Infect. Ther.* 10, 701–706
5. Kumar, G., Kumar, N., Taneja, A., Kaleekal, T., Tarima, S., McGinley, E., Jimenez, E., Mohan, A., Khan, R. A., Whittle, J., Jacobs, E., and Nanchal, R. (2011) Nationwide trends of severe sepsis in the 21st century (2000-2007). *Chest* 140, 1223–1231
6. Ludwig, K. R., and Hummon, A. B. (2017) Mass spectrometry for the discovery of biomarkers of sepsis. *Mol. Biosyst.* 13, 648–664
7. Geyer, P. E., Holdt, L. M., Teupser, D., and Mann, M. (2017) Revisiting biomarker discovery by plasma proteomics. *Mol. Syst. Biol.* 13, 942
8. Geyer, P. E., Voytik, E., Treit, P. V., Doll, S., Kleinhempel, A., Niu, L., Müller, J. B., Buchholtz, M., Bader, J. M., Teupser, D., Holdt, L. M., and Mann, M. (2019) Plasma Proteome Profiling to detect and avoid sample-related biases in biomarker studies. *EMBO Mol. Med.* 11,
9. Geyer, P. E., Wewer Albrechtsen, N. J., Tyanova, S., Grassl, N., Iepsen, E. W., Lundgren, J., Madsbad, S., Holst, J. J., Torekov, S. S., and Mann, M. (2016) Proteomics reveals the effects of sustained weight loss on the human plasma proteome. *Mol. Syst. Biol.* 12, 901
10. Namas, R., Zamora, R., Namas, R., An, G., Doyle, J., Dick, T. E., Jacono, F. J., Androulakis, I. P., Nieman, G. F., Chang, S., Billiar, T. R., Kellum, J. A., Angus, D. C., and Vodovotz, Y. (2012) Sepsis: Something old, something new, and a systems view. *J. Crit. Care* 27, 314.e1-314.e11
11. Cao, Z., Yende, S., Kellum, J. A., Angus, D. C., and Robinson, R. A. S. (2014) Proteomics reveals age-related differences in the host immune response to sepsis. *J.*

12. Qian, W. J., Monroe, M. E., Liu, T., Jacobs, J. M., Anderson, G. A., Shen, Y., Moore, R. J., Anderson, D. J., Zhang, R., Calvano, S. E., Lowry, S. F., Xiao, W., Moldawer, L. L., Davis, R. W., Tompkins, R. G., Camp, D. G., Smith, R. D., Baker, H. V., Bankey, P., Billiar, T. R., Brownstein, B. H., Chaudry, I. H., Cobb, J. P., Fay, A., Feezor, R. J., Freeman, B., Gamelli, R. L., Gibran, N. S., Harbrecht, B. G., Hayden, D., Herndon, D. N., Horton, J. W., Hunt, J. L., Johnson, J. L., Laudanski, K., Lederer, J. A., Logvinenko, T., Maier, R. V., Mannick, J. A., McKinley, B., Miller-Graziano, C. L., Minei, J. P., Mindrinos, M., Moore, E. E., Moore, F. A., Nathens, A. B., O'Keefe, G. E., Rahme, L. G., Remick, D. G., Schoenfeld, D., Shapiro, M. B., Sheridan, R. L., Silver, G. M., Somers, S., Toner, M., Warren, H. S., West, M. A., Wolf, S. E., Yarmush, M., and Young, V. R. (2005) Quantitative proteome analysis of human plasma following in vivo lypopolysaccharide administration using ¹⁶O/¹⁸O labeling and the accurate mass and time tag approach. *Mol. Cell. Proteomics* 4, 700–709
13. Shen, Z., Want, E. J., Chen, W., Keating, W., Nussbaumer, W., Moore, R., Gentle, T. M., and Siuzdak, G. (2006) Sepsis plasma protein profiling with immunodepletion, three-dimensional liquid chromatography tandem mass spectrometry, and spectrum counting. *J. Proteome Res.* 5, 3154–3160
14. Garcia-Obregon, S., Azkargorta, M., Seijas, I., Pilar-Orive, J., Borrego, F., Elortza, F., Boyano, M. D., and Astigarraga, I. (2018) Identification of a panel of serum protein markers in early stage of sepsis and its validation in a cohort of patients. *J. Microbiol. Immunol. Infect.* 51, 465–472
15. Thavarajah, T., dos Santos, C. C., Slutsky, A. S., Marshall, J. C., Bowden, P., Romaschin, A., and Marshall, J. G. (2020) The plasma peptides of sepsis. *Clin. Proteomics* 17, 26
16. Pepys, M. B., and Baltz, M. L. (1983) Acute Phase Proteins with Special Reference to C-Reactive Protein and Related Proteins (Pentaxins) and Serum Amyloid A Protein. *Adv. Immunol.* 34, 141–212
17. Pepys, M. B., and Hirschfield, G. M. (2003) C-reactive protein: a critical update. *J. Clin. Invest.* 111, 1805–1812
18. Póvoa, P., Coelho, L., Almeida, E., Fernandes, A., Mealha, R., Moreira, P., and Sabino, H. (2005) C-reactive protein as a marker of infection in critically ill patients. *Clin. Microbiol. Infect.* 11, 101–108

19. Ryoo, S. M., Han, K. S., Ahn, S., Shin, T. G., Hwang, S. Y., Chung, S. P., Hwang, Y. J., Park, Y. S., Jo, Y. H., Chang, H. L., Suh, G. J., You, K. M., Kang, G. H., Choi, S. H., Lim, T. H., Kim, W. Y., Kim, Y. J., Lee, S. W., Jo, I. J., Kim, M. J., Kwon, W. Y., Lee, H. J., Shin, J. H., and Ko, B. S. (2019) The usefulness of C-reactive protein and procalcitonin to predict prognosis in septic shock patients: A multicenter prospective registry-based observational study. *Sci. Rep.* 9, 1–8
20. Reily, C., Stewart, T. J., Renfrow, M. B., and Novak, J. (2019) Glycosylation in health and disease. *Nat. Rev. Nephrol.* 15, 346–366
21. Stowell, S. R., Ju, T., and Cummings, R. D. (2015) Protein glycosylation in cancer. *Annu Rev Pathol* 10, 473–510
22. Pinho, S. S., and Reis, C. A. (2015) Glycosylation in cancer: Mechanisms and clinical implications. *Nat. Rev. Cancer* 15, 540–555
23. Varki A, Cummings RD, Esko JD, et al. (2009) *Essentials of Glycobiology. 2nd edition. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2009. Part II, Structure and Biosynthesis.*
24. Lauc, G., Essafi, A., Huffman, J. E., Hayward, C., Knežević, A., Kattla, J. J., Polašek, O., Gornik, O., Vitart, V., Abrahams, J. L., Pučić, M., Novokmet, M., Redžić, I., Campbell, S., Wild, S. H., Borovečki, F., Wang, W., Kolčić, I., Zgaga, L., Gyllensten, U., Wilson, J. F., Wright, A. F., Hastie, N. D., Campbell, H., Rudd, P. M., and Rudan, I. (2010) Genomics Meets Glycomics—The First GWAS Study of Human N-Glycome Identifies HNF1 α as a Master Regulator of Plasma Protein Fucosylation. *PLoS Genet.* 6, e1001256
25. Klarić, L., Tsepilov, Y. A., Stanton, C. M., Mangino, M., Sikka, T. T., Esko, T., Pakhomov, E., Salo, P., Deelen, J., McGurnaghan, S. J., Keser, T., Vučković, F., Ugrina, I., Krištić, J., Gudelj, I., Štambuk, J., Plomp, R., Pučić-Baković, M., Pavić, T., Vilaj, M., Trbojević-Akmačić, I., Drake, C., Dobrinić, P., Mlinarec, J., Jelušić, B., Richmond, A., Timofeeva, M., Grishchenko, A. K., Dmitrieva, J., Bermingham, M. L., Sharapov, S. Z., Farrington, S. M., Theodoratou, E., Uh, H. W., Beekman, M., Slagboom, E. P., Louis, E., Georges, M., Wuhrer, M., Colhoun, H. M., Dunlop, M. G., Perola, M., Fischer, K., Polasek, O., Campbell, H., Rudan, I., Wilson, J. F., Zoldoš, V., Vitart, V., Spector, T., Aulchenko, Y. S., Lauc, G., and Hayward, C. (2020) Glycosylation of immunoglobulin G is regulated by a large network of genes pleiotropic with inflammatory diseases. *Sci. Adv.* 6, 24
26. Lin, Y.-H., Zhu, J., Meijer, A. B., Vojtech, F., and Heck, A. J. R. (2019)

Glycoproteogenomics: a frequent gene polymorphism affects the glycosylation pattern of the human serum fetuin/ α -2-HS-glycoprotein. *Mol. Cell. Proteomics*, mcp.RA119.001411

27. Tamara, S., Franc, V., and Heck, A. J. R. (2020) A wealth of genotype-specific proteoforms fine-tunes hemoglobin scavenging by haptoglobin. *Proc. Natl. Acad. Sci. U. S. A.* 117, 15554–15564
28. De Coux, A., Tian, Y., Deleon-Pennell, K. Y., Nguyen, N. T., De Castro Brás, L. E., Flynn, E. R., Cannon, P. L., Griswold, M. E., Jin, Y. F., Puskarich, M. A., Jones, A. E., and Lindsey, M. L. (2015) Plasma glycoproteomics reveals sepsis outcomes linked to distinct proteins in common pathways. *Crit. Care Med.* 43, 2049–2058
29. Cichy, J., Potempa, J., Chawla, R. K., and Travis, J. (1995) Stimulatory effect of inflammatory cytokines on α 1-antichymotrypsin expression in human lung-derived epithelial cells. *J. Clin. Invest.* 95, 2729–2733
30. Aronsen, K. F., Ekelund, G., Kindmark, C., and Laurell, C. B. (1972) Sequential changes of plasma proteins after surgical trauma. *Scand. J. Clin. Lab. Invest.* 29, 127–136
31. Heimburger, N., and Haupt, H. (1965) Charakterisierung von α 1 X-glykoprotein als chymotrypsin-Inhibitor des humanplasmas. *Clin. Chim. Acta* 12, 116–118
32. RubinS, H., mei Wang, Z., Nickbargg, E. B., McLarneyS, S., NaidooQ, N., Schoenbergg, O. L., JohnsonQll, J. L., and Cooperman, B. S. (1990) Cloning, Expression, Purification, and Biological Activity of Recombinant Native and Variant Human arl-Antichymotrypsins*. *J. OP Biol. Chem.* 265, 1199–1207
33. York, W. S., Mazumder, R., Ranzinger, R., Edwards, N., Kahsay, R., Aoki-Kinoshita, K. F., Campbell, M. P., Cummings, R. D., Feizi, T., Martin, M., Natale, D. A., Packer, N. H., Woods, R. J., Agarwal, G., Arpinar, S., Bhat, S., Blake, J., Castro, L. J. G., Fochtman, B., Gildersleeve, J., Goldman, R., Holmes, X., Jain, V., Kulkarni, S., Mahadik, R., Mehta, A., Mousavi, R., Nakarakommula, S., Navelkar, R., Pattabiraman, N., Pierce, M. J., Ross, K., Vasudev, P., Vora, J., Williamson, T., and Zhang, W. (2020) GlyGen: Computational and Informatics Resources for Glycoscience. *Glycobiology* 30, 72–73
34. Kuscuoğlu, D., Janciauskiene, S., Hamesch, K., Haybaeck, J., Trautwein, C., and Strnad, P. (2018) Liver- master and servant of serum proteome. *J. Hepatol.* 69, 512–524
35. Vandevyver, S., Dejager, L., Vandenbroucke, R. E., and Libert, C. (2014) An acute

phase protein ready to go therapeutic for sepsis. *EMBO Mol. Med.* 6, 2–3

36. Langhans, C., Weber-Carstens, S., Schmidt, F., Hamati, J., Kny, M., Zhu, X., Wollersheim, T., Koch, S., Krebs, M., Schulz, H., Lodka, D., Saar, K., Labeit, S., Spies, C., Hubner, N., Spranger, J., Spuler, S., Boschmann, M., Dittmar, G., Butler-Browne, G., Mouly, V., and Fielitz, J. (2014) Inflammation-Induced Acute Phase Response in Skeletal Muscle and Critical Illness Myopathy. *PLoS One* 9, e92048
37. Wu, D., Struwe, W. B., Harvey, D. J., Ferguson, M. A. J., and Robinson, C. V (2018) N-glycan microheterogeneity regulates interactions of plasma proteins. *Proc. Natl. Acad. Sci. U. S. A.* 115, 8763–8768
38. Yang, Y., Franc, V., and Heck, A. J. R. (2017) Glycoproteomics: A Balance between High-Throughput and In-Depth Analysis. *Trends Biotechnol.* 35, 598–609
39. Rosati, S., Rose, R. J., Thompson, N. J., Van Duijn, E., Damoc, E., Denisov, E., Makarov, A., and Heck, A. J. R. (2012) Exploring an orbitrap analyzer for the characterization of intact antibodies by native mass spectrometry. *Angew. Chemie - Int. Ed.* 51, 12992–12996
40. Rosati, S., Yang, Y., Barendregt, A., and R Heck, A. J. (2014) Detailed mass analysis of structural heterogeneity in monoclonal antibodies using native mass spectrometry. *Nat. Protoc.* 9,
41. Bern, M., Caval, T., Kil, Y. J., Tang, W., Becker, C., Carlson, E., Kletter, D., Sen, K. I., Galy, N., Hagemans, D., Franc, V., and Heck, A. J. R. (2018) Parsimonious Charge Deconvolution for Native Mass Spectrometry. *J. Proteome Res.* 17, 1216–1226
42. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 26, 1367–1372
43. Dellinger, R. P., Levy, M. M., Rhodes, A., Annane, D., Gerlach, H., Opal, S. M., Sevransky, J. E., Sprung, C. L., Douglas, I. S., Jaeschke, R., Osborn, T. M., Nunnally, M. E., Townsend, S. R., Reinhart, K., Kleinpell, R. M., Angus, D. C., Deutschman, C. S., Machado, F. R., Rubenfeld, G. D., Webb, S. A., Beale, R. J., Vincent, J.-L., and Moreno, R. (2013) Surviving Sepsis Campaign. *Crit. Care Med.* 41, 580–637
44. Hotchkiss, R. S., Monneret, G., and Payen, D. (2013) Immunosuppression in sepsis: A novel understanding of the disorder and a new therapeutic approach. *Lancet Infect. Dis.* 13, 260–268

45. Sharma, N. K., and Salomao, R. (2017) Sepsis Through the Eyes of Proteomics. *SHOCK* 47, 17–25
46. Hayashi, N., Yamaguchi, S., Rodenburg, F., Ying Wong, S., Ujimoto, K., Miki, T., and Iba, T. (2019) Multiple biomarkers of sepsis identified by novel time-lapse proteomics of patient serum. *PLoS One* 14, e0222403
47. Pimienta, G., Heithoff, D. M., Rosa-Campos, A., Tran, M., Esko, J. D., Mahan, M. J., Marth, J. D., and Smith, J. W. (2019) Plasma Proteome Signature of Sepsis: a Functionally Connected Protein Network. *Proteomics* 19, 1800389
48. Lissauer, M. E., Johnson, S. B., Siuzdak, G., Bochicchio, G., Whiteford, C., Nussbaumer, B., Moore, R., and Scalea, T. M. (2007) Coagulation and complement protein differences between septic and uninfected systemic inflammatory response syndrome patients. *J. Trauma - Inj. Infect. Crit. Care* 62, 1082–1092
49. Sharma, N. K., Tashima, A. K., Brunialti, M. K. C., Ferreira, E. R., Torquato, R. J. S., Mortara, R. A., MacHado, F. R., Assuncao, M., Rigato, O., and Salomao, R. (2017) Proteomic study revealed cellular assembly and lipid metabolism dysregulation in sepsis secondary to community-acquired pneumonia. *Sci. Rep.* 7, 1–13
50. Chen, K. F., Chaou, C. H., Jiang, J. Y., Yu, H. W., Meng, Y. H., Tang, W. C., and Wu, C. C. (2016) Diagnostic accuracy of lipopolysaccharide-binding protein as biomarker for sepsis in adult patients: A systematic review and meta-analysis. *PLoS One* 11,
51. Zhang, Y., Zhang, J., Sheng, H., Li, H., and Wang, R. (2019) in *Advances in Clinical Chemistry* (Academic Press Inc.), pp 25–80.
52. Clerc, F., Reiding, K. R., Jansen, B. C., Kammeijer, G. S. M., Bondt, A., and Wuhler, M. (2015) Human plasma protein N-glycosylation. *Glycoconj. J.*, 1–35
53. Čaval, T., Tian, W., Yang, Z., Clausen, H., and Heck, A. J. R. (2018) Direct quality control of glycoengineered erythropoietin variants. *Nat. Commun.* 9, 3342
54. Wohlschlagel, T., Scheffler, K., Forstenlehner, I. C., Skala, W., Senn, S., Damoc, E., Holzmann, J., and Huber, C. G. (2018) Native mass spectrometry combined with enzymatic dissection unravels glycoform heterogeneity of biopharmaceuticals. *Nat. Commun.* 9, 1–9
55. Lin, Y. H., Zhu, J., Meijer, S., Franc, V., and Heck, A. J. R. (2019) Glycoproteogenomics: A frequent gene polymorphism affects the glycosylation pattern of the human serum fetuin/α-2-HS-Glycoprotein. *Mol. Cell. Proteomics* 18, 1479–1490

56. Hollander, C., Westin, U., Wallmark, A., Piitulainen, E., Sveger, T., and Janciauskiene, S. M. (2007) Plasma levels of alpha1-antichymotrypsin and secretory leukocyte proteinase inhibitor in healthy and chronic obstructive pulmonary disease (COPD) subjects with and without severe α 1-antitrypsin deficiency. *BMC Pulm. Med.* 7, 1–12
57. POS, O., STELT, M. E., WOLBINK, G.-J., NUSTEN, M. W. N., TEMPEL, G. L., and DIJK, W. (1990) Changes in the serum concentration and the glycosylation of human α 1-acid glycoprotein and α 1-protease inhibitor in severely burned persons: relation to interleukin-6 levels. *Clin. Exp. Immunol.* 82, 579–582
58. De Graaf, T. W., Van der Stelt, M. E., Anbergen, M. G., and van Dijk, W. (1993) Inflammation-induced expression of sialyl lewis X-containing glycan structures on α 1-Acid glycoprotein (orosomucoid) in human sera. *J. Exp. Med.* 177, 657–666
59. Higai, K., Aoki, Y., Azuma, Y., and Matsumoto, K. (2005) Glycosylation of site-specific glycans of α 1-acid glycoprotein and alterations in acute and chronic inflammation. *Biochim. Biophys. Acta - Gen. Subj.* 1725, 128–135
60. Higai, K., Miyazaki, N., Azuma, Y., and Matsumoto, K. (2008) Transcriptional regulation of the fucosyltransferase VI gene in hepatocellular carcinoma cells. *Glycoconj. J.* 25, 225–235
61. Aird, W. C. (2003) The role of the endothelium in severe sepsis and multiple organ dysfunction syndrome. *Blood* 101, 3765–3777
62. Shapiro, N. I., Schuetz, P., Yano, K., Sorasaki, M., Parikh, S. M., Jones, A. E., Trzeciak, S., Ngo, L., and Aird, W. C. (2010) The association of endothelial cell signaling, severity of illness, and organ dysfunction in sepsis. *Crit. Care* 14, R182
63. Gornik, O., and Lauc, G. (2008) Glycosylation of serum proteins in inflammatory diseases. *Dis. Markers* 25, 267–278
64. Yin, H., An, M., So, P. kin, Wong, M. Y. M., Lubman, D. M., and Yao, Z. (2018) The analysis of alpha-1-antitrypsin glycosylation with direct LC-MS/MS. *Electrophoresis* 39, 2351–2361
65. McCarthy, C., Saldova, R., Wormald, M. R., Rudd, P. M., McElvaney, N. G., and Reeves, E. P. (2014) The role and importance of glycosylation of acute phase proteins with focus on alpha-1 antitrypsin in acute and chronic inflammatory conditions. *J. Proteome Res.* 13, 3131–3143
66. Shental-Bechor, D., and Levy, Y. (2009) Folding of glycoproteins: toward

understanding the biophysics of the glycosylation code. *Curr. Opin. Struct. Biol.* 19, 524–533

67. Varki, A. (2017) Biological roles of glycans. *Glycobiology* 27, 3–49
68. LAINE, A., HACHULLA, E., STRECKER, G., MICHALSKI, J.-C., and WIERUSZESKI, J.-M. (1991) Structure determination of the glycans of human-serum alpha1-antichymotrypsin using ¹H-NMR spectroscopy and deglycosylation by N-glycanase. *Eur. J. Biochem.* 197, 209–215

Supporting information for Chapter 4

Supplementary Figures

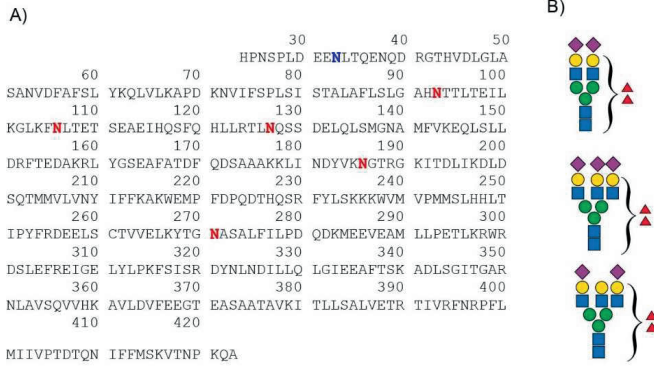


Figure S1. The amino acid sequence of AACT with reported N-glycosylation sites. **A)** amino acid sequence of AACT with highly occupied N-glycosylation sites highlighted in red, and low occupancy site highlighted in blue(31). **B)** Most abundant AACT glycoforms reported in literature(67).

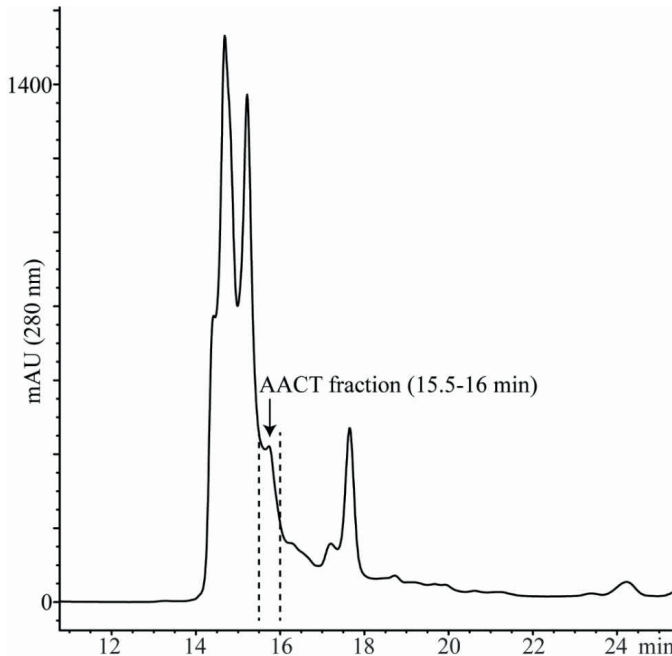


Figure S2. Ion Exchange Chromatogram. An example of a plasma sample ion-exchange chromatogram is shown, and the collected fraction containing AACT is highlighted.

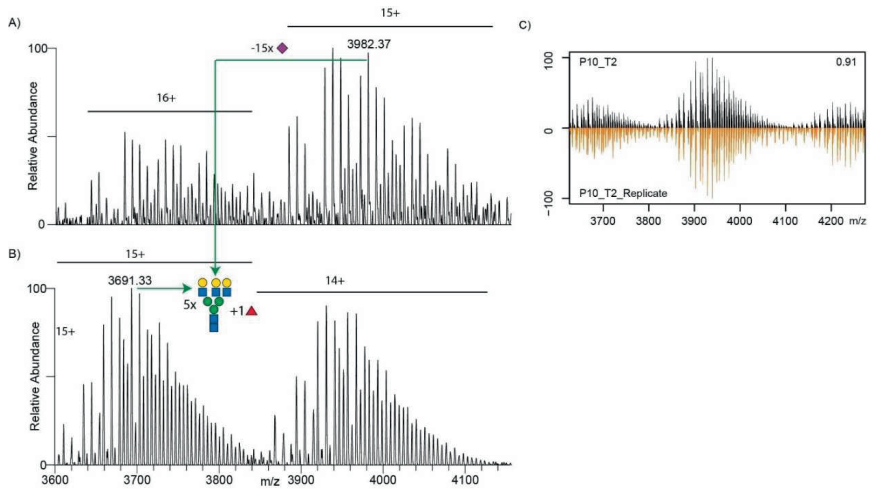


Figure S3. Clarification of glycoproteoform profile of AACT by sialidase treatment. A) Example of sialylated AACT and sialidase treated AACT in **B)**. Horizontal lines denote charge state and green arrow connects composition corresponding to three triantennary N-glycans and a fucose with 15 sialic acids. Loss of 15 sialic acids can be observed in sialidase treated spectrum. **C)** Technical reproducibility of AACT measurement, on top T2 of P10 is shown and below in orange a technical replicate of the same patient and timepoint is shown. Upper right corner shows a Person correlation score.

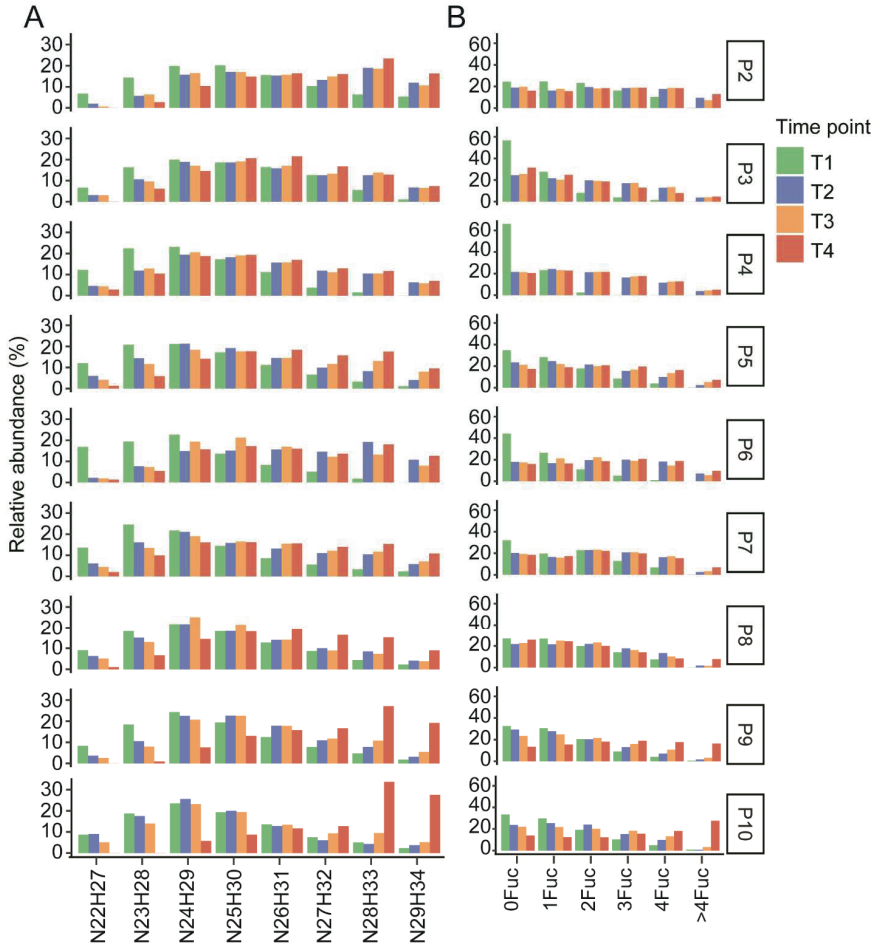


Figure S4. Changes in the level of branching and fucosylation of AACT across four time points for each patient. A) Glycoprotoform compositions with the same HexNAcHex (N=HexNAc, H=Hex) composition were summed up together and plotted for all four time points B) All glycoprotoform compositions containing at least one fucose were summed up and sorted on their fucose content

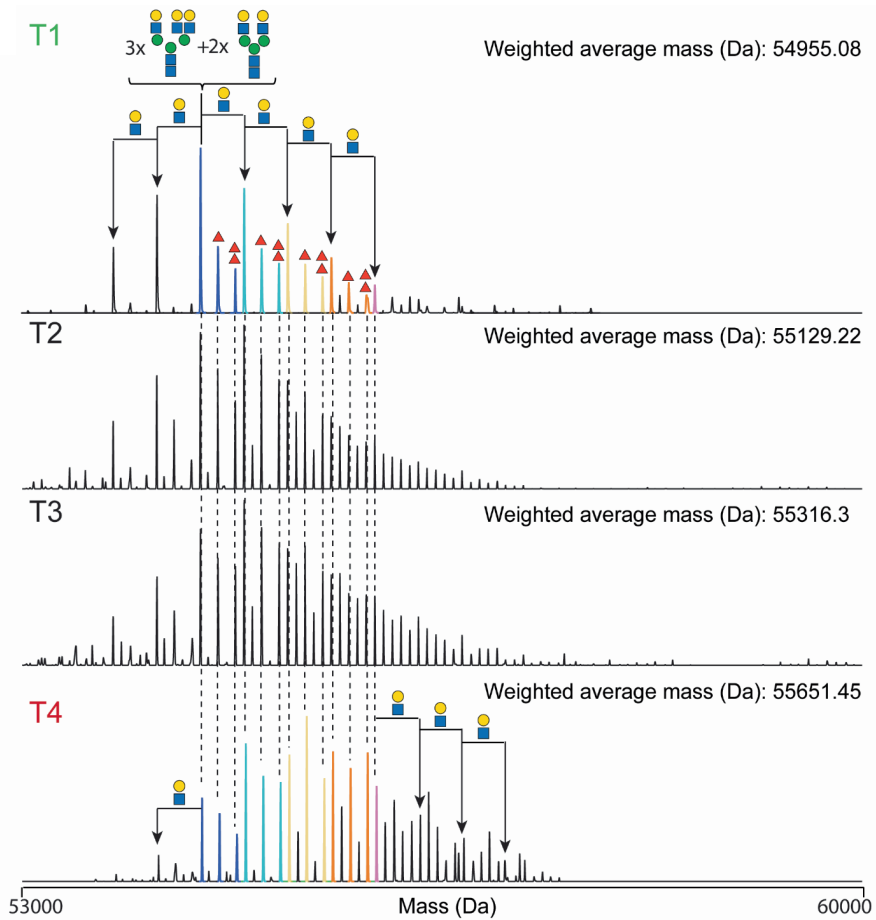


Figure S5. Illustrative deconvoluted native mass spectra displaying the continuous remodeling of alpha-1-antichymotrypsin glycosylation across the four time points as here depicted for patient P1.

Supplementary Table

Sample	Abundance of CRP (mg/l)
P1T1	71
P1T2	375
P1T3	451
P1T4	78
P2T1	121
P2T2	375
P2T3	333
P2T4	91
P3T1	73
P3T2	137
P3T3	71
P3T4	52
P4T1	73
P4T2	220
P4T3	180
P4T4	87
P5T1	78
P5T2	282
P5T3	202
P5T4	57
P6T1	48
P6T2	440
P6T3	463
P6T4	126
P7T1	94
P7T2	340
P7T3	314
P7T4	10
P8T1	71
P8T2	332
P8T3	302
P8T4	13
P9T1	54
P9T2	386
P9T3	318
P9T4	134
P10T1	84
P10T2	228
P10T3	228
P10T4	96

Supplementary Table S3-Clinical data on the plasma abundance level of the C-reactive protein (CRP)

Sample	Average mass
P1T1	54955.08
P1T2	55129.22
P1T3	55316.30
P1T4	55651.45
P2T1	55453.59
P2T2	55982.52
P2T3	55963.03
P2T4	56214.17
P3T1	55225.81
P3T2	55671.06
P3T3	55717.17
P3T4	55797.07
P4T1	54819.75
P4T2	55602.88
P4T3	55619.43
P4T4	55717.09
P5T1	55045.27
P5T2	55445.24
P5T3	55683.63
P5T4	55954.02
P6T1	54800.60
P6T2	55953.61
P6T3	55792.84
P6T4	56008.99
P7T1	55051.12
P7T2	55568.76
P7T3	55676.97
P7T4	55876.14
P8T1	55218.53
P8T2	55464.65
P8T3	55446.61
P8T4	55873.27
P9T1	55206.44
P9T2	55485.60
P9T3	55646.12
P9T4	56354.12
P10T1	55245.46
P10T2	55301.04
P10T3	55520.53
P10T4	56623.58

Supplementary Table S4--Calculated average mass of AACT (Da)

The following supplementary tables are not printed because they are oversized.

Supplementary Table S1: Proteome search and label-free quantification result.

Supplementary Table S2: Normalized proteome label-free quantification result in each sample to T1.

Supplementary Data: The list containing relative abundances of AACT proteoforms in all samples.

Chapter 5

A quantitative longitudinal inventory of the *N*-glycoproteome of human milk from a single donor reveals the highly variable repertoire and dynamic site-specific changes

Jing Zhu^{a,b,c,d}, Yu-Hsien Lin^{a,b,d}, Kelly A. Dingess^{a,b}, Marko Mank^d, Bernd Stahl^d and Albert J. R. Heck^{a,b}

^aBiomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Padualaan 8, 3584 CH Utrecht, The Netherlands

^bNetherlands Proteomics Center, Padualaan 8, 3584 CH Utrecht, The Netherlands

^cBeijing Institute of Nutritional Resources, 100069, Beijing, China

^dDanone Nutricia Research, Uppsalalaan 12, 3584 CT Utrecht, The Netherlands

^eEqual Contribution

* Authors for Correspondence: Albert Heck, a.j.r.heck@uu.nl (+31302536797)

Based on:

Zhu J, Lin YH, Dingess KA, Mank M, Stahl B, Heck AJR. Quantitative Longitudinal Inventory of the N-Glycoproteome of Human Milk from a Single Donor Reveals the Highly Variable Repertoire and Dynamic Site-Specific Changes. *J Proteome Res.* 2020;19(5):1941-1952. doi:10.1021/acs.jproteome.9b00753

Abbreviations

ACN: Acetonitrile

AGC: Automated gain control

CAA: Chloroacetamide

ET_hcD: Electron-transfer/higher-energy collision dissociation

FA: Formic acid

FDR: False discovery rate

HCD: Higher energy c-trap dissociation

HILIC: Hydrophilic-interaction chromatography

HMOs: Human milk oligosaccharides

HR/AM: High resolution/accurate mass

IMAC: Immobilized metal ion affinity chromatography

MRM: Multiple reaction monitoring

MS: Mass spectrometry

MS/MS: Tandem mass spectrometry

pIgR: Polymeric immunoglobulin receptor

PRM: Parallel reaction monitoring

PRTC: Pierce peptide retention time calibration

PSM: Peptide to spectrum matches

PTM: Post-translational modification

RT: Retention time

SDC: Sodium deoxycholate

sIgA: secretory immunoglobulin A

SIM: Selective ion monitoring

TCEP: Tris(2-carboxyethyl)phosphine hydrochloride

TFA: Trifluoroacetic acid

Abstract

Protein *N*-glycosylation on human milk proteins assists in protecting the infant's health and functions amongst others as competitive inhibitors of pathogen binding and immunomodulators. Due to the individual uniqueness of each mother's milk and the overall complexity and temporal changes of protein *N*-glycosylation, analysis of the human milk *N*-glycoproteome requires longitudinal personalized approaches, providing protein- and *N*-site-specific quantitative information. Here we describe an automated platform using HILIC-based cartridges enabling the proteome-wide monitoring of intact *N*-glycopeptides using just a digest of 150 µg of breast milk protein. We were able to map around 1700 glycopeptides from 110 glycoproteins covering 191 glycosites, of which 43 sites have not been previously reported with experimental evidence. We next quantified 287 of these glycopeptides originating from 50 glycoproteins using a targeted proteomics approach. Although each glycoprotein, *N*-glycosylation site and attached glycan revealed distinct dynamic changes, we did observe a few general trends. For instance, fucosylation, especially terminal fucosylation, increased across the lactation period. Building on the improved glycoproteomic approach outlined above, future studies are warranted to reveal the potential impact of observed glycosylation microheterogeneity on the healthy development of infants.

Keywords: human milk / glycoprotein / milk proteins / glycopeptides / HILIC-based platform / mass spectrometry / *N*-glycosylation / *N*-glycopeptide enrichment / targeted *N*-glycopeptide quantification/ lactation

Introduction

Human milk is a remarkable multifunctional fluid of the mammary gland, providing a varying amount of nutritional and non-nutritional, bioactive components to satisfy the specific demand of the newborn¹. A major component of human milk are glycans, occurring both in free form as lactose and oligosaccharides², and as well as conjugated to glycoproteins primarily in the form of *N*- and/or *O*-linked glycosylation³. Protein glycosylation is of special interest in milk due to the widespread functional capacity. For instance, glycosylated proteins are relevant to proteolytic susceptibility⁴, as competitive inhibitors of pathogen binding^{5, 6} and immunomodulators⁵, all together working to protect the infant's health⁷. Like other components in human milk, protein glycosylation is quite diverse among individual mothers and it thought to also vary quite dynamically across lactation⁸. To understand the bioactive effects of the continuing changes of glycosylation in human milk, it is necessary that protein glycosylation analysis is done both in a personalized manner and longitudinally.

The characterization of human milk glycosylation is challenging due to the large dynamic range of human milk proteins⁹ and the complexity of protein glycosylation. The first challenge is posed by the co-occurrence of high-abundant non-glycosylated proteins in human milk, such as beta-casein that accounts for ~30% of the proteins in weight,⁹ which diminishes the detectability of glycopeptides in a standard proteomics approach. Furthermore, the microheterogeneity introduced by harboring different types of glycans on the same protein glycosylation site further confounds the complexity of the sample and additionally the specific analysis of glycopeptides^{10, 11}.

To date, the glycopeptide-centric analyses of human milk protein glycosylation is still rather in its infancy. Thus, further developing this analysis is essential for obtaining information on site-specific glycosylation. Recent studies mainly focused on the mapping of *N*-glycosylation sites of enzymatic deglycosylated peptides¹²⁻¹⁴ or on enzymatically released *N*-glycans^{15, 16}. To name a few, a recent study from Cao et al¹² investigated human colostrum and mature milk samples, and reported 68 and 58 *N*-glycoproteins and 111 and 96 *N*-glycosites, respectively, using a lectin enrichment approach and subsequent identification of the deglycosylated peptides. Picariello et al¹⁴ used hydrophilic-interaction chromatography (HILIC) glycopeptide enrichment and applied this to a milk sample collected on day 7 after parturition, also merely analyzing the deglycosylated peptides, identifying 63 *N*-glycosylated sites originating from 32 glycoproteins. Released *N*-glycans have revealed that a specific feature of human milk glycosylation is the high degree of abundant fucosylation, when compared with bovine milk¹⁵ or healthy human serum¹⁷. For instance, Nwosu et al¹⁵ compared human and bovine milk and found substantial differences in their fucosylated *N*-glycans. They reported that 75% of the *N*-

glycans in human milk correspond to fucosylated glycans while this number was only 31% in bovine milk. Moreover, human milk exhibited more glycoforms harboring multiple fucoses when compared to bovine milk. About 60% of the fucosylated *N*-glycans in human milk were reported to be multi-fucosylated, whereas this number was just 30% in bovine milk. Also Dallas et al¹⁶ quantified the abundances of *N*-glycans in human milk and found that 84% of the total were fucosylated *N*-glycans. They further reported that fucosylated *N*-glycans in human milk harbor either 1 (36%), 2 (32%), 3 (14%) or more (2%) fucose moieties. This high degree of fucosylation in human milk is in sharp contrast to what has been observed in human serum. For instance, Yoshida et al¹⁷ reported that the majority of *N*-glycans in human serum were glycans without fucose (around 800 pmol/ μ l), followed by glycans with one fucose (around 300 pmol/ μ l) and glycans with two fucoses (only around 0.6 pmol/ μ l).

Here, we report on a high-throughput method to profile the human milk *N*-glycoproteome, taking a glycopeptide centric approach, with the aim to monitor changes occurring over the lactation period in a single individual donor. Therefore, we collected human milk samples at 9 time points covering the early, transitional and mature lactational stages. We enriched for intact *N*-glycopeptides using an automated HILIC-based setup, and identified the *N*-glycopeptides by higher energy C-trap dissociation (HCD) triggered electron-transfer/higher-energy collision dissociation (EThcD). We also quantified many *N*-glycopeptides using scheduled single ion monitoring (SIM) and parallel reaction monitoring (PRM). Overall our dataset provides a comprehensive view at the human milk *N*-glycoproteome. We identify and profile 191 glycosites on 110 glycoproteins, of which 43 sites on 32 proteins can be considered novel. Moreover these 191 sites harbor in total 1697 different glycans, i.e. on average 8 per site. We could also quantify 287 glycopeptides originating from 50 glycoproteins across the lactational period, which allowed us to observe distinctive site-specific glycosylation profiles. From this comprehensive profiling, the picture emerges that the human milk glycoproteome is rather complex and dynamic, likely reflecting the changing demands of the newborns growing and developing immune system and gut microbiota.

Experimental Procedures

Human Milk Sample Collection

Human milk samples were collected longitudinally from one healthy donor in week 1, 2, 3, 4, 6, 8, 10, 12 and 16 (for a total of 9 time-points). Written informed consent was obtained prior to the collection of any samples. All samples used were donated to Danone Nutricia Research in accordance to the Helsinki Declaration II. Milk samples were collected from one breast, as a complete breast expression, between the hours of 9-11 AM, taking the breast that had not

been used for feeding within a 2-hour window. Immediately after pumping, samples were transferred to a 2 mL Eppendorf tube containing protease inhibitors (Complete Protease Inhibitor Cocktail Tablets from Roche) and were then frozen at -20°C. Samples were transferred to the laboratory on dry ice and stored at -80°C until thawed for analysis.

Experimental Design and Statistical Rationale

For all milk samples, three technical replicate analyses were performed. The mean \pm standard deviation of the three technical replicates was calculated and the proportion was calculated by using values of the mean of three technical replicates.

Chemicals and Materials

Unless otherwise specified, all chemicals and reagents were obtained from Sigma-Aldrich (Steinheim, Germany). Lys-C was obtained from Wako (Neuss, Germany). The Oasis PRiME HLB plate was purchased from Waters (Etten-Leur, the Netherlands). Formic acid (FA) was purchased from Merck (Darmstadt, Germany). Acetonitrile (ACN) was purchased from Biosolve (Valkenswaard, The Netherlands). The Pierce peptide retention time calibration (PRTC) mixture was obtained from Thermo Fisher Scientific (Bremen, Germany). Milli-Q was produced by an in-house system (Millipore, Billerica, MA).

Milk Serum Separation and Protein Digestion

For each time-point, an aliquot of 200 μ L of a thawed whole milk sample was centrifuged at 1000 \times g for 1 h at 4 °C whereafter the upper-fat layer was discarded. The lower skimmed milk was further ultracentrifuged at 150,000 \times g for 1 h at 4 °C to remove the insoluble pellet and the supernatant milk serum was retained. The protein concentration was estimated by measuring absorbance at 280nm on a Nanodrop Spectrophotometer (Nanodrop 2000, Thermo Scientific). Up to a final concentration of 1% w/v sodium deoxycholate (SDC), 100mM Tris-HCl (pH 8.5), 5mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) and 30 mM chloroacetamide (CAA) were added to the milk serum. Trypsin and Lys-C were added to a 1:50 and 1:100 ratio (w/w), respectively. Digestion was performed overnight at 37 °C. The next day SDC was removed *via* acid precipitation (0.5% trifluoroacetic acid) (TFA) and the final peptide concentration was estimated by measuring the absorbance at 280 nm on a Nanodrop Spectrophotometer (Nanodrop 2000, Thermo Scientific)¹⁸. The peptides were desalted by using an Oasis PRiME HLB plate then dried and stored at -80 °C.

Automated HILIC-based glycopeptide enrichment

The HILIC-based glycopeptide enrichment was performed automatically with triplicates for each milk serum sample using an AssayMAP Bravo robot (Agilent technologies) coupled with HILIC-based cartridges (GlykoPrep® APTS Cleanup Module, ProZyme, CA). The HILIC-based columns were first washed with 50 µL of 1% FA and equilibrated with 50 µL of loading buffer (80% ACN/0.5% TFA). The peptide digests (150 µg) were reconstituted with 50 µL of loading buffer and loaded onto the column. The cartridges were washed with 50 µL of loading buffer, and the glycopeptides were step eluted with 75% ACN/0.5% TFA, 70% ACN/0.5% TFA, 65% ACN/0.5% TFA and 0.5% FA. These samples were dried down and stored at -80 °C until subjected to LC-MS/MS.

Full Proteome Analysis of Human Milk

To estimate protein abundances in the human milk proteome, 800 ng of the mixture of non-enriched tryptic peptides were analyzed using an Agilent 1290 Infinity HPLC system (Agilent Technologies, Waldbronn, Germany) coupled on-line to a Q Exactive HF mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). The peptides were first trapped by using a 100 µm inner diameter 2 cm trap column (in-house packed with ReproSil-Pur C18-AQ, 3 µm) (Dr. Maisch GmbH, Ammerbuch-Entringen, Germany) coupled to a 50 m inner diameter 50 cm analytical column (in-house packed with Poroshell 120 EC-C18, 2.7 µm) (Agilent Technologies, Amstelveen, The Netherlands). The mobile-phase solvent A consisted of 0.1% FA in water, and the mobile-phase solvent B consisted of 0.1% FA in ACN. Trapping was performed at a flow rate of 5 µl/min for 5 min with 0% B and peptides were eluted using a passively split flow of 300 nl/min for 170 min with 10% to 36% B over 155 min, 36% to 100% B over 3 min, 100% B for 1 min, 100% to 0% B over 1 min, and finally held at 0% B for 10 min. Peptides were ionized using a spray voltage of 1.9 kV and a heated capillary. The mass spectrometer was set to acquire full-scan MS spectra (375-1600 m/z) for a maximum injection time of 20 ms at a mass resolution of 60,000 and an automated gain control (AGC) target value of 3e6. Up to 15 of the most intense precursor ions were selected for tandem mass spectrometry (MS/MS). HCD MS/MS (200-2000 m/z) acquisition was performed in the HCD cell, with the readout in the Orbitrap mass analyzer at a resolution of 30,000 (isolation window of 1.4 Th) and an AGC target value of 1e5 or a maximum injection time of 50 ms with a normalized collision energy of 27%.

Raw shotgun LC-MS/MS data files were searched using a processing workflow in Proteome Discoverer (version 2.2, Thermo Scientific) using the Mascot search engine (version 2.6.1) against the Swiss-Prot database (release date: Feb 2017, 20 172 entries, taxonomy: Homo

sapiens) using fixed Cys carbamidomethylation and variable Met oxidation of peptides as search variables. Trypsin was chosen for cleavage specificity with a maximum of two missed cleavages allowed. The searches were performed using a precursor mass tolerance of 10 ppm and a fragment mass tolerance of 0.05 Da followed by data filtering using Percolator, resulting in a 1% false discovery rate (FDR). Only peptide to spectrum matches (PSMs) with Mascot score > 20 were accepted. The full proteome search result was then used as a focused database for the identification of glycopeptides (1259 entries). For label-free quantification, the node called minora feature detector was used with high PSM confidence, a minimum of 5 non-zero points in a chromatographic trace, a minimum number of 2 isotopes and maximum retention time difference of 0.2 min for isotope peaks. The consensus workflow in Proteome Discoverer was used to open the search results and enable retention time (RT) alignment with a maximum RT shift of 5 min and mass tolerance of 10 ppm in order to match the precursor between runs. Label-free quantification of peptides was performed using the intensities of the extracted ion chromatogram (XIC). Protein intensity was determined by the average intensity of unique peptides. Protein abundance was estimated by taking the proportion of the protein intensity for each protein to the total protein intensity. To relatively compare the proportional change in each time point, the protein abundance in week 1 was defined as standard, giving the adjustment factor of 1. Other adjustment factors of proteins in other weeks were calculated using their relative abundances normalized to that measured value for that protein in week 1.

Result

Identification of Human Milk Glycopeptides

All peptides from the automated HILIC-based glycopeptide enrichment were separated and analyzed using the same HPLC system as used for the global proteome analysis, albeit now coupled on-line to an Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) using a 90 min gradient, as follows: 0-5 min, 100% solvent A; 13-44% solvent B for 65 min; 44-100% solvent B for 5 min; 100% solvent B for 5 min; 100% solvent A for 15 min. Peptides were ionized using a 2.0 kV spray voltage. For the MS scan, the mass range was set from 350-1800 m/z for a maximum injection time of 50 ms at a mass resolution of 60,000 and an AGC target value of $4e5$ in the Orbitrap mass analyzer. The dynamic exclusion was set to 30 s for an exclusion window of 10 ppm with a cycle time of 3 s. Charge-states screening was enabled, and precursors with 2+ to 8+ charge states and intensities > $1e5$ were selected for tandem mass spectrometry (MS/MS). HCD MS/MS (120-2100 m/z) acquisition was performed in the HCD cell, with the readout in the Orbitrap mass analyzer at a resolution of 30,000 (isolation window of 1.6 Th) and an AGC target value of $5e4$ or a maximum injection time of 75 ms with a normalized collision energy of 30%. If at least 2 out of 3 oxonium ions of

glycopeptides (138.0545+, 204.0687+ or 366.1396+) were observed, EThcD MS/MS on the same precursor was triggered (isolation window of 1.6 Th) and fragment ions (120–2100 m/z) were analyzed in the Orbitrap mass analyzer at a resolution of 30,000, AGC target value of $2e5$ or a maximum injection time of 250 ms with activation of ETD and supplemental activation with a normalized collision energy of 27%. To obtain MS/MS scans from more various precursors, we made reference samples which mixed glycopeptides from week 1, 4, 10 (reference 1), week 2, 6, 12 (reference 2) and week 3, 8, 16 (reference 3) and measured each reference sample three times using the same LC-MS/MS method for glycopeptide identification except that we applied various dynamic exclusion times; 30 s, 60 s and 180 s, respectively.

All the raw files obtained for the glycopeptide identification including those of the reference samples were processed in Proteome Discoverer (version 2.2, Thermo Scientific) using the Byonic node (Protein Metrics Inc, ver 3.2.0) searching against a targeted milk protein database (1259 entries) based on our own data from the global milk proteome analysis using the following search parameters: trypsin digestion with a maximum of 2 missed cleavages, precursor ion mass tolerance, 10 ppm; fragmentation type, both HCD & EThcD; fragment mass tolerance, 20 ppm; carbamidomethylation of cysteines as a fixed modification; variable modifications: methionine oxidation. For glycan analysis we used a Byonic database of 182 glycans with no multiple fucoses, whereby we added manually several reported glycan compositions with multiple fucoses (Supplementary Table S1). The maximum number of precursors per scan was set to one and the FDR as 1%. Only PSMs matched to EThcD spectra with non-negligible error probabilities $|\log \text{Prob}| > 4.0$ and Byonic score ≥ 200 were accepted. MS1 feature of high confident PSMs was captured by the node called minora feature detector and precursor was matched between runs by enabling retention time (RT) alignment with a maximum RT shift of 5 min and mass tolerance of 10 ppm.

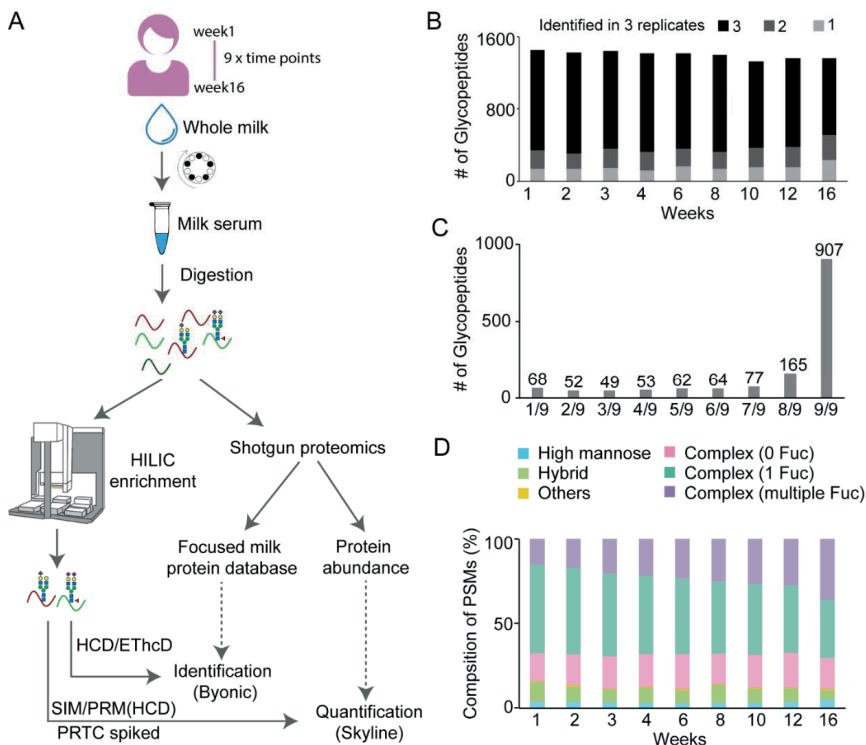


FIG. 1. Experimental overview of the human milk glycopeptide-centric proteome analysis.

A. Overview of experimental steps taken. Human milk samples were collected longitudinally from the same individual across the early, transitional and mature lactational stages, at week 1, 2, 3, 4, 6, 8, 10, 12 and 16. By centrifugation and ultracentrifugation, the separated milk serum was subsequently tryptic digested, whereafter the total peptide concentration was determined. Milk proteins and their abundances across lactation were measured by standard shotgun proteomics approaches. *N*-glycopeptides present in human milk were enriched by using HILIC implemented on an automated AssayMAP Bravo platform. The resulting glycopeptides were then analyzed in HCD triggered EThcD and searched by Byonic. A dedicated human milk protein database was used based on the proteomics data from the shotgun identification. *N*-glycopeptides were next quantified by a targeted assay with scheduled SIM and PRM with spiked PRTC peptides to supervise the variability between runs and retention time shifts. These data were quantitatively analyzed using Skyline. The intensity of the glycopeptide was normalized to the intensity of its originating protein, as measured in the shotgun quantification.

B. Total numbers of *N*-glycopeptides found in 1, 2 or 3 technical replicates. About 90% of the glycopeptides were identified in at least two replicates at all time-points. Week 16 revealed somewhat lower reproducibility, albeit still 83%.

C. Total number of *N*-glycopeptides detected in at least 2 out of 3 technical replicates (1497 glycopeptides) found at all 9 timepoints. Only 4.5% (68) were solely found in one time-point, the majority 60.6% (907) were detected at all time-points, indicating the ability to reproducibly enrich from the different milk samples.

D. General composition of the observed glycoforms, based on detected glycopeptide PSMs. The observed glycoforms were grouped into high mannose, hybrid and complex types, whereby the complex type was further divided into subgroups without fucose, with one fucose and with multiple fucoses, and others. See Supplementary Table S5 for details about the glycan identifiers. The composition of complex with one fucose was observed to gradually decrease while the

composition of complex with multiple fucoses was found to relatively increase from early to mature lactation.

Quantification of Human Milk Glycopeptides

First, we built a library in Skyline¹⁹ (Skyline-daily, version 4.2.1.18305) based on the m/z and retention time of each identified glycopeptide precursor and its HCD fragmentation pattern. Since the used version of Skyline was only able to accommodate *b* and *y* ions, we manually added oxonium ions and Y ions (peptide backbone ions carrying a glycan fragment from the glycosidic bond cleavage). The targeted peptides were selected based on their chromatographic trace and intensity. Due to the lack of appropriate stable isotope-labeled glycopeptide standards, we spiked PRTC peptides equally in all the samples, which helped to monitor potential retention time shifts and variability between MS runs. Around 10% of the enriched glycopeptides and 50 fmol of PRTC mixture were analyzed by LC-MS/MS with the same HPLC system and mass spectrometer as described above for the identification of the glycopeptides, except for MS and MS/MS scans were acquired in SIM and PRM mode, respectively. For MS scans, the m/z range was set from 350-2000 with a resolution of 30,000 using an AGC setting of 5e4, maximum IT of 54 ms, one microscan, a 1.6 Th isolation window, 27% normalized collision energy, and 3 min retention time window. For MS/MS scans, the m/z range was set from 120-3000 with a resolution of 30,000 using an AGC setting of 5e4, maximum IT of 54 ms, one microscan, a 1.6 Th isolation window, 30% normalized collision energy, and 3 min retention time window.

Further data analysis was done in Skyline as well. Skyline initially scanned for up to 5 isotopes of the precursors and all *b* and *y* fragment ions, oxonium ions and Y ions for each peptide. Each peak was manually assessed, where after ions with interferences were removed. The retention time and rank of fragment ions were checked. If the fragment ions co-eluted and the rank were the same as recorded in the above-mentioned library, the chromatographic peak of the precursor was undoubtedly pinpointed and the abundance of the glycopeptides was represented by the max intensity of the top 3 isotopes of the precursors. The abundances of the PRTC peptides were extracted to determine the coefficient of variation (CV). To relatively quantify the abundance of each glycopeptide over the lactation period, the abundance was normalized using the glycoprotein abundances obtained in our global human milk proteome analysis of the same samples.

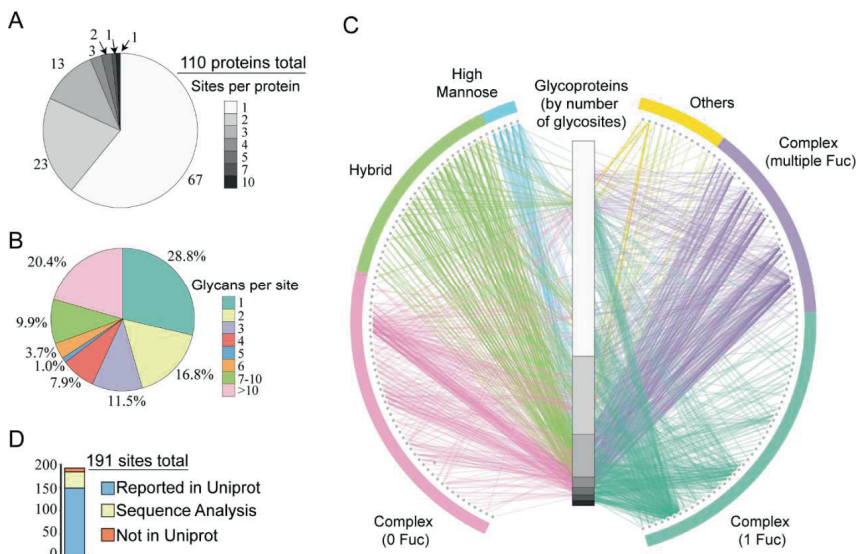


FIG. 2. Characteristics of the glycoproteins, glycopeptides and glycosites observed in the human milk *N*-glycoproteome. A. Distribution of the number of glycosites observed on each of the 110 identified glycoproteins. B. Distribution of the number of different glycans observed on each glycosite. C. Glycoprotein-glycan network map displaying what sort of glycans (outer circle, 160 total) modify which proteins (inner bar, 110 total). Glycoproteins are sorted as in A, by the number of observed glycosites (note same scale as the distribution in A). The observed glycans were therefore organized using a standard classification, and edges are colored by the glycan node from which they originate. See Supplementary Table S5 for more details about the glycan identifiers. D. Approximately 22.5% (43) of the identified glycosites are not yet annotated in UniProt. Of these 35 could be hypothesized to be genuine *N*-glycosylation sites based on the presence of the N-X-S/T sequon, whereas 8 were not annotated at all.

Results and Discussion

Human Milk Sample Collection and Experimental Workflow

Here, we aimed to extensively characterize the temporal changes in the *N*-glycoproteome of a single mother's breast milk during the course of a lactation period. The overall experimental workflow is depicted in Figure 1A and further detailed in the Experimental Procedures. In brief, we collected milk samples longitudinally from the same mother across early, transitional and mature lactational stages, *i.e.* at week 1, 2, 3, 4, 6, 8, 10, 12 and 16. Following ultracentrifugation, the obtained milk serum was digested by LysC and trypsin, whereafter the total peptide concentration was estimated. First, abundances of the human milk proteins across the lactation period were determined by label-free standard shotgun proteomics approaches. In parallel, *N*-glycopeptides present in human milk were selectively enriched using HILIC-based

cartridges implemented on an automated 96-well plate AssayMAP Bravo robot. The resulting glycopeptides were subsequently analyzed by LC-MS/MS, using HCD triggered EThcD as the fragmentation method, whereafter the spectra were searched and identified using the Bionic node in Proteome Discoverer. A dedicated human milk protein database was made based on the 1259 proteins identified in the shotgun proteomics experiments. A selected set of human milk *N*-glycopeptides were quantified in more detail using a targeted assay with scheduled SIM and PRM, using a spiked-in PRTC peptide standard, to correct for the variability between runs and potential LC retention time shifts. These data were further quantitatively analyzed using Skyline. The intensity of each glycopeptide was normalized to the intensity of its protein of origin, as measured in the shotgun quantification of the same human milk sample.

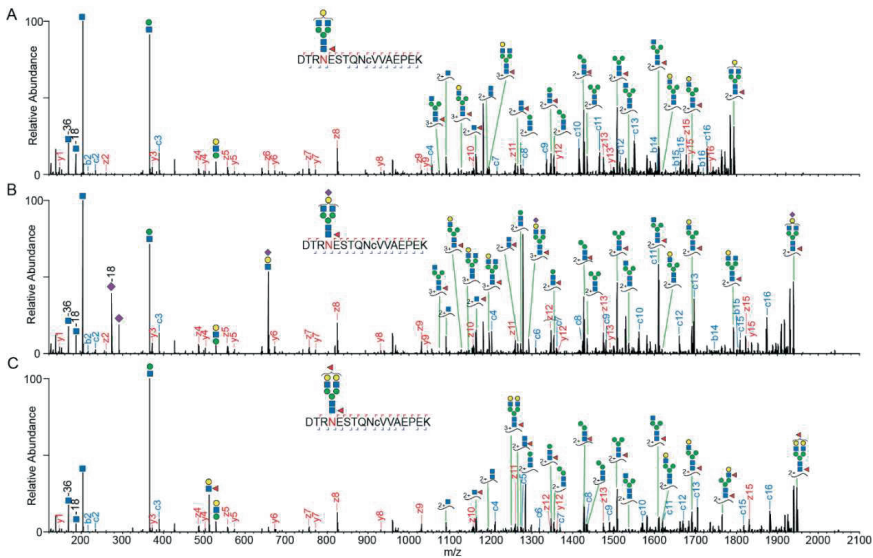


FIG. 3. Illustrative EThcD spectra of glycopeptides derived from alpha-S1-casein on a novel *N*-glycosylation site. Three examples of the same glycopeptide; DTRNESTQNCVVAEPEK of alpha-S1-casein (66-82), harboring various glycoforms. The EThcD fragmentation of these glycopeptides resulted in full sequence coverage and the fragment ion (Y1f) of the amino acid backbone with HexNAc(1)Fuc(1) indicated one core fucose. A, EThcD fragmentation spectra of the glycopeptide harboring a HexNAc(4)Hex(4)Fuc(1). B, EThcD fragmentation spectra of the glycopeptide harboring a HexNAc(4)Hex(4)Fuc(1)Sia(1). C, EThcD fragmentation spectra of the glycopeptide harboring a HexNAc(4)Hex(5)Fuc(2). Lower case c in the peptide sequence indicates a carbamylated Cysteine.

Mapping the Human Milk *N*-Glycoproteome across Lactation

To retain only the most confident glycopeptide identifications, we only accepted PSMs matched to EThcD spectra with strict thresholds, namely a non-negligible error probability $|\log$

Prob $>$ 4.0 and a Bionic score \geq 200. In general a high Bionic score improves accuracy but at the expense of lowering the number of IDs. We selected cut-off filters based on previous work by Lee *et al.*²⁰ who investigated the relationship between the Bionic score and the accuracy of the IDs, whereby it was concluded that a Bionic score $>$ 200 was a stringent but good cut-off. For the Log probability cut-off we selected a value $>$ 4, as a value of $>$ 1 results in an estimated 0.33% FDR for glyco PSMs²¹. After applying this stringent cut-off, in total 22,614 PSMs could be identified (Supplementary Table S2), resulting in 1697 unique intact *N*-glycopeptides (Supplementary Table S3) originating from 110 different human milk glycoproteins. The identified glycoproteins spanned a range in abundance very similar to those observed in the full human milk proteome analysis (Supplementary Figure 1). As a quality check, we next assessed the reproducibility in the automated enrichment of glycopeptides among all technical replicates (Figure 1B). About 90% of all glycopeptides were identified in at least 2 out of 3 technical replicates at all time-points, whereby only week 16 revealed a somewhat lower reproducibility, albeit still around 83%, revealing the robustness and reproducibility of our automated glycopeptide enrichment.

We next inspected the 1497 glycopeptides which were found in at least 2 out of 3 technical replicates among the 9 different time-points (Figure 1C). Only 4.5% (68) of these were uniquely found in one time-point, the majority 61% (907) were detected at all time points. Thus, qualitatively the glycopeptide variety is rather constant over the lactation period. Following this, we evaluated the nature of the glycans identified based on detected PSMs (Figure 1D). The observed glycoforms were grouped into high mannose, hybrid and complex types, whereby the complex type was further divided into subgroups without fucose, with one fucose and with multiple fucoses. The detailed compositions of the glycans can be found in Supplementary Table S5. The majority of the identified glycans (67-70% of the PSMs) at each time point consisted of complex *N*-glycans with none, one or multiple fucoses. This is in line with previous results for human milk analysis focused on released *N*-glycans^{15, 16}. Among the complex *N*-glycans, we noticed that the glycans carrying one fucose gradually decreased while those harboring multiple fucose moieties increased in time from early to mature lactation, again consistent with previous findings based on studies focused on released *N*-glycans²².

Having established a global picture of the glycan distribution as occurring on glycoproteins in human milk, we used this data to make a rough comparison to the global distribution of glycans in human serum, taking for serum data reported earlier²³ (Supplementary Fig 2). This comparison revealed, not surprisingly, that global glycoproteome compositions in human milk and serum are quite distinct. In particular, hybrid structures and complex glycans harboring multiple fucoses seems to be relatively more abundant in human milk *versus* serum.

Since we detected the majority of glycopeptides at all investigated time-points, we combined all data to get a comprehensive picture of the human milk *N*-glycoproteome. We first counted how many glycosites we identified for each glycoprotein (Figure 2A). The number of glycosites per glycoprotein varied from 1 to 10 (Supplementary Table S4). In total, 191 glycosites were identified on cumulatively 110 glycoproteins. These proteins together harbor 546 *N*-glycosylation sites annotated in Uniprot by either experimental evidence or sequon prediction. Of all glycoproteins, 67 were detected with one glycosite, 23 with two glycosites, 13 with three glycosites, 3 with four glycosites, 2 with five glycosites, 1 with seven glycosites and 1 with ten glycosites. The latter is the glycoprotein Tenascin (TNC), which is known to be involved in the protection against viral infections, signifying the importance of glycans in host defense²⁴. We next investigated the degree of glycan microheterogeneity observed on each of the 191 glycosites (Figure 2B, Supplementary Table S4). Overall, 72% of the glycosites were identified with more than one glycoform, 20% of the glycosites were even harboring more than 10 glycoforms. This clearly indicates the high degree of glycan microheterogeneity observed on human milk glycoproteins. To further visualize the data we adapted a glycoprotein-glycan network diagram²¹ (Figure 2C), in which the variety of glycans (outer nodes, 160 in total, Supplementary Table S5) are connected to their glycoproteins of origin (inner column, organized by the number of glycosites, 110 in total). Again the most eye-catching feature is the high prevalence of complex glycans harboring multiple fucosylation. This network diagram also indicates that the complex glycans are found on nearly all glycoproteins, while the majority of high mannose and hybrid glycans occurred mostly on glycoproteins harboring multiple glycosylation sites.

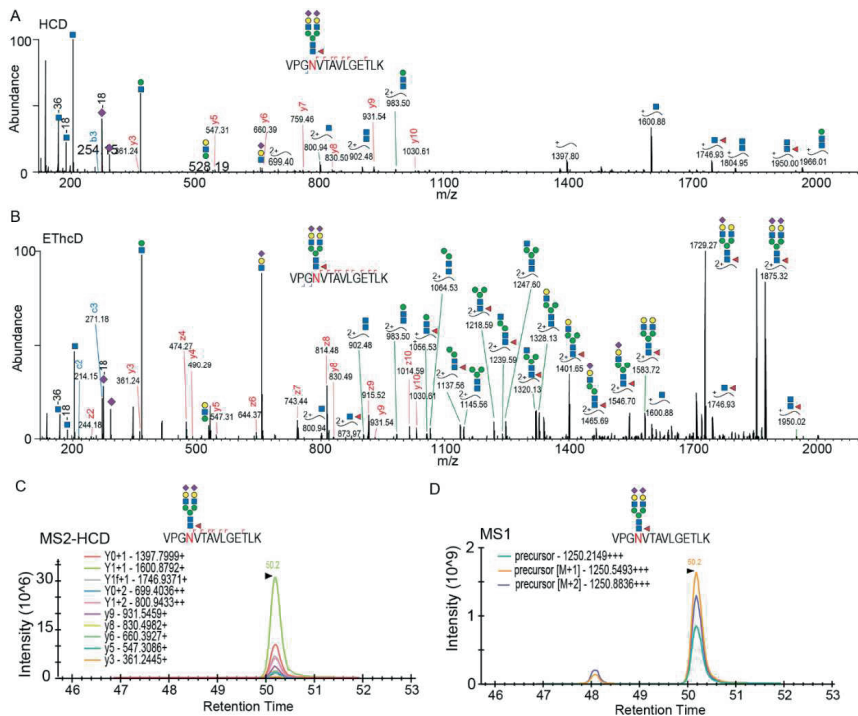


FIG.4. Targeted quantitative monitoring of glycopeptides. For illustrative purposes a glycopeptide of the polymeric immunoglobulin receptor; VPGNVTAVLGETLK is taken, harboring a HexNAc(4)Hex(5)Fuc(1)Sia(2) moiety. A. HCD spectrum of the glycopeptide precursor (m/z 1250.2149+++). B. EThcD spectrum of the same precursor which improved the identification and site-assignment. C. Elution profile of Y and y ions which coeluted and whose rank was matched with the HCD spectrum in A. D. The chromatographic peak of the precursor could be pinpointed by the MS2-HCD elution profile, whereby the max intensity of the top 3 isotopes was used to determine the abundance of the glycopeptide of interest.

Identification of novel N-glycosylation sites

We evaluated the N-glycosylation sites found in our data *versus* those annotated in UniProt (Figure 2D). Most of the sites we report here (149 out of 191 glycosites) are annotated with experimental evidence in Uniprot, however, we detected 43 potentially novel glycosites all displaying the well-known N-X-S/T sequon. In UniProt these 43 were annotated as putative sites without any experimental evidence (35 glycosites) or not annotated at all (8 glycosites). Especially for these putative novel sites we not only took the stringent cut-off in Byonic, but also manually validated the MS/MS spectra using the following procedure. 1) All these spectra should contain at least two Y ions (peptide backbone ion carrying a glycan fragment from the glycosidic bond cleavage) and two oxonium ions, which strongly indicate that the spectra

originated from intact glycopeptides. 2) The ETHcD fragmentation ions should contain the *c* or *z* ion pairs with a difference in mass of the intact glycan and corresponding amino acids, which is essential for an unambiguous site localization. Of all the novel sites identified the most surprising one was Asn69 (sequon: NES) in alpha-S1-casein, since this protein is quite abundant and a very well-studied milk highly phosphorylated protein²⁵. We generated very strong evidence for Asn69 *N*-glycosylation, observing many different glycopeptides harboring this site. Due to a variable number of miscleavages we identified EKQTDEIKDTRN₆₉ESTQNCVVAEPEK (58-82), QTDEIKDTRN₆₉ESTQNCVVAEPEK (60-82), DTRN₆₉ESTQNCVVAEPEK (66-82), DTRN₆₉ESTQNCVVAEPEKMESSISSSSEEMSLSK (66-98) and N₆₉ESTQNCVVAEPEK (69-82). Among these, the backbone sequence of DTRN₆₉ESTQNCVVAEPEK (66-82) was found to be detected with most PSMs and glycoform variants. The corresponding ETHcD spectra showed both glycan and backbone fragmentation, enabling confident assignments of the glycoforms and glycosylation site (Figure 3). Moreover, the glycopeptides with two fucoses could be assigned as one core fucosylation and one terminal fucosylation by the Y1f ion (peptide backbone ion carrying a glycan fragment HexNAc(1)Fuc(1) from the glycosidic bond cleavage) and HexNAc(1)Hex(1)Fuc(1) oxonium ion (*m/z* 512.197+).

Additionally, we investigated the site occupancy of this site by directly monitoring in the milk serum shotgun experiments (i.e. without any enrichment or depletion) for the DTRNESTQNCVVAEPEK peptide in its unmodified form and for the same peptide harboring the glycoform HexNAc(4)Hex(5)Fuc(1)Sia(1). We selected this glycopeptide as it was on average about 15 times more intense than all the others (Supplementary Figure 3A). The peak area of extracted ion chromatograms of the unmodified peptide in different time-points showed gradually decrease over lactation and was almost not detected in week 16, while the glycopeptide harboring the glycoform HexNAc(4)Hex(5)Fuc(1)Sia(1) declined less (Supplementary Figure 3B). By comparing the ion intensities of the most abundant glycopeptide and the corresponding non-modified peptide we extracted that the site-occupancy of Asn69 in alpha-S1-casein ranged from 20% at the early stages to almost fully occupied (98%) at later stages of the lactation (Supplementary Figure 3C). This substantial change in occupancy may indicate that the site has a functional relevance. However, uncomfortable at the late stages we almost did not detect the non-modified peptide anymore, opening up the question of whether possibly other modifications may play a role as well. Alpha-S1-casein is best known as a highly phosphorylated protein harboring close to a dozen phosphorylation sites. Among these phosphorylation sites, Ser71, which was recently reported to be phosphorylated in breast milk²⁶, happens to be in the sequon of the *N*-glycosylation we found (NES), and thus on the peptide stretch we monitored. This opens the question of whether a potential positive or negative

crosstalk between *N*-glycosylation and phosphorylation may occur. In human alpha-S1-casein has 4 isoforms produced by alternative splicing²⁷ (Supplementary Figure 4). Interestingly, in isoform 3, both Asn69 and Ser71 are deleted. Moreover, aligning alpha-S1-casein protein sequences across different species (Supplementary Figure 4), revealed that only the human sequence carries this *N*-glycosylation site.



FIG. 5. Site-specific glycosylation varies across stages of lactation, as illustrated for pIgR.

The temporal changes in abundance of the protein polymeric immunoglobulin receptor (pIgR) and the glycans it harbors are depicted. A. Overview of the pIgR protein sequence, the identified glycosites, the glycan compositions and the abundances over the 9 measured time-points. The protein sequence chains are color-coded, wherein light blue and blue represent the signal peptide and the mature protein, respectively. Indigo dots inserted in the protein sequence with an asterisk indicate the UniProt-annotated *N*-glycosylation sites with evidence. A red circle indicates that the glycosite is identified in this study. B. Taking the value in week 1 as the reference point, the abundance of the protein pIgR gradually decreases marginally over the 16 weeks of monitoring its abundance. In contrast, the glycan distribution on N469 and N499 (taken by the summed intensities of biantennary glycopeptides VPGN₄₆₉VTAVLGETLK, WN₄₉₉NTGcQALPSQDEGSPK, respectively) show a diverging behavior with a 2-2.5-fold increase in abundance after 6-8 weeks, which decreases back to the abundance observed at 1 week at week 16. Micro-heterogeneity in glycan composition monitored over the lactation period as observed for B. VPGN₄₆₉VTAVLGETLK and C. WN₄₉₉NTGcQALPSQDEGSPK. For pIgR, the glycoforms harboring fucose moieties displayed a relatively increase over the lactation period. In general, the site-specific glycosylation patterns were quite distinctive for each glycoprotein and each *N*-glycosite.

Targeted Monitoring of Intact Glycopeptides across Lactation

Observing that the glycan occupancy of certain *N*-glycopeptides varies substantially during the lactation period, we decided to monitor a subset of observed glycopeptides by a targeted proteomics assay. Therefore, we first built a library with relevant *m/z* and retention times of selected intact glycopeptide precursors and their corresponding HCD fragmentation pattern. The glycopeptide identification based on the HCD spectrum (Figure 4A) was confirmed by their matching EThcD spectra (Figure 4B). We then scheduled, for all selected glycopeptides, the precursor and its fragmentation ions to be measured. In a 90 min LC-MS gradient, we did target and measure 287 glycopeptides originating from 50 glycoproteins by selected ion monitoring (SIM) and parallel reaction monitoring (PRM) respectively, catching at least 10 data points in their chromatographic peaks (Supplementary Figure 5A and Supplementary Figure 5B). In the MS/MS-HCD spectra, Y ions are normally quite abundant²⁸. We could use these ions in addition to the *b* and *y* ions to check if they were co-eluted and compare the rank to that recorded in the above-mentioned library (Figure 4C). If so, we could undoubtedly pinpoint the chromatographic peak of the precursor and use the maximum intensity of the top 3 isotopes to represent the abundance of the glycopeptides (Figure 4D). We also noticed the separation of glycopeptide isoforms in MS1 XICs, however, we could often not directly distinguish their exact form. Therefore, in our analysis, we took the most intense MS1 XIC of the detected isoforms. For some glycopeptides containing Met, we noticed variable oxidation occurring, thus we summed intensity of the oxidation (+/-) forms. Considering the lack of commercial C¹³-labeled glycopeptide standards, we spiked in the PRTC mixture equally in all samples to monitor potential LC retention time shifts and variability between MS runs. We could calculate the CV of the intensities of PRTC peptides as a measure of variation. Of all PRTC peptides, 70% had CVs less than 20%, another 30% had CV less than 40% (Supplementary Figure 5C). Based on this observation, we concluded that the variability between runs was within a reasonable range. Based on our in parallel acquired proteome data, we noticed that the abundances of certain proteins also showed changes across the lactation period. Thus, to assess the total quantitative change in specific *N*-glycoform abundances, the raw intensity obtained from the targeted monitoring was normalized to the in parallel measured protein abundances, using the protein abundances from week 1 as the baseline (Supplementary Table S6).

Visual Summary for Each Glycoprotein and Each Glycosite

To better illustrate the dynamics observed in the site-specific glycosylation of all detected glycoproteins, we mapped on the UniProt sequence of each glycoprotein the detected glycans

per site with their relative abundances over the lactation period (using log₁₀ of average normalized intensities) as shown in Supplementary Figure 6 (Uniprot Accession). The diversity of glycan microheterogeneity can be exemplified by comparing different glycoproteins. Some proteins have only one glycosite that harbors a great variety of glycans, e.g., alpha-S1-casein (Supplementary Figure 6. P47710), and immunoglobulin J chain (Supplementary Figure 6. P01591). However, other glycoproteins have several glycosylation sites, with limited variability in the glycans modifying them, e.g., hemopexin (Supplementary Figure 6. P02790). Also some glycosites exhibit higher variability in glycans than other sites on the same protein, e.g., lactoferrin (Supplementary Figure 6. P02788). Interestingly, for lactoferrin, the glycosite with lower glycan heterogeneity (Asn642) is the site which is unique to human lactoferrin, other sites (Asn156, Asn497) are more conserved glycosylation sites.

We choose to further demonstrate the variability of glycans and glycosites by focusing on one example, namely the polymeric immunoglobulin receptor (pIgR) protein, which has seven potential *N*-glycosylation sites. The protein pIgR in human milk is quite abundant²⁹ and plays an essential role in the formation and secretion of secretory immunoglobulin A (sIgA), which is the dominant immunoglobulin in human milk and essential for infant health³⁰. Not only is pIgR essential for the secretion of IgA, it has known antimicrobial functionality, and can also be secreted on its own providing protection to sIgA from degradation³¹. We identified a stunning number of 201 glycopeptides in total covering all seven potential glycosylation sites of pIgR, whereby each site, displayed a striking different degree of heterogeneity (Figure 5A). We quantified 34 of the glycopeptides over 9 time-points, further elucidating quantitative differences. Among all detected glycosites, Asn469 and Asn499 accounted for the most PSMs and the highest degree of heterogeneity. We zoomed into the variation in occupancy of these two sites by different glycans, especially biantennary complex glycans with or without fucose moieties across lactation. Firstly, as shown in Figure 5B the pIgR protein does not change that much in abundance over the monitored lactation period, although it does show a gradual decrease by up to about 20% at the latest time point. After normalization for protein level variation, the glycopeptides harboring Asn469 and Asn499 in pIgR did reveal much more variability and site-specific changes over the lactation period. . We next compared the longitudinal variations in the different glycans occupying these two sites. In Figure 5C it is shown that for VPGN₄₆₉VTAVLGETLK, glycoforms harboring one fucose declined in abundance, while glycoforms with multiple fucoses increased especially HexNAc(4)Hex(5)Fuc(2)NeuAc(1) from 36% to 58%. For WN₄₉₉NTGCQALPSQDEGPSK (Figure 5D), the glycoforms harboring one fucose increased while those without fucose decreased. Notably, glycan HexNAc(4)Hex(5)Fuc(1) increased from 5% to 18% and HexNAc(4)Hex(5)Fuc(1)NeuAc(1) from 16% to 39%, whereas HexNAc(4)Hex(5)NeuAc(1)

dropped from 51% to 31% and HexNAc(4)Hex(5)NeuAc(2) from 18% to 5%. Our data are also supported by earlier studies which reported that fucosylation increases over lactation^{32, 33}. For instance, Landberg et al.³² observed that with the progression of lactation, the levels of fucosylation on *N*-glycans of the bile-salt-stimulated lipase protein increased. Related to our work, Barboza et al.³³ analyzed the released *N*-glycans of human milk lactoferrin from day 1 to 72 postpartum and found decreased levels of mono-fucosylation and increased levels of multiple fucosylation. They also showed that enzymatic removal of the fucose moieties from lactoferrin affected the ability of bacteria to bind to epithelial cells, whereas lactoferrin with fucose attached significantly inhibited pathogen adhesion. In our data, we observed similar trends for the glycopeptides of lactoferrin (Supplementary Figure 6 P02788).

In mammals, core fucosylation, which is fucosylation on the GlcNAc linked to asparagine in the core of *N*-glycans, is the most common type of fucose modification³⁴. Thus mono-fucosylation tends to be core fucosylation, while additional fucosylation of *N*-glycans is likely terminal fucosylation, for which we also find evidence in our EThcD spectra. The function of multiple fucosylation in the course of lactation, especially the site-specific effects are still understudied. Nevertheless, the feature of having terminal fucosylation on glycoproteins represents a structural homology to human milk oligosaccharides (HMOs), which may lead to a similar function of the fucosylated HMOs. There is evidence from studies investigating HMOs that have shown that HMOs increase in complexity and diversity, especially regarding fucosylation, throughout lactation to meet the needs of the increasing diversity of the gut microbiota³⁵⁻³⁷. This could be an explanation for the changes observed in our study regarding the glycoproteome, especially with regard to increasing fucosylation observed from several different glycoproteins and varying glycosites across lactation. Potentially then these observed differences are related to the changing needs of the infant to meet functional demands for growth and development and of diversifying gut microbiome. *N*-glycans could be effectively released by the bacterium-derived glycohydrolases, e.g. Endo- β -*N*-acetylglucosaminidases³⁸, and serve as the selective growth substrates for infant-associated gut microbes^{39, 40}. Additionally, the attached *N*-glycans could slow down the proteolytic digestion in infants, which is conceivably supported by evidence that many glycoproteins like lactoferrin⁴¹ and sIgA⁴² are not fully absorbed in the small intestine, but rather can be found back as intact protein in infant stool samples, facilitating that these proteins serve many important roles in pathogen defense and regulation of cellular proliferation and differentiation⁵.

Conclusions

In this study we characterized the human milk *N*-glycoproteome of a single donor over 9 time points during the lactation period, characterizing 191 glycosites on cumulatively 110 human milk glycoproteins. In total we were able to assign 1697 glycopeptides, indicating that many sites were occupied by a wide variety of glycans. In more quantitative detail we performed targeted proteomics on 287 glycopeptides originating from 50 glycoproteins monitoring their abundance over the lactational period, through which we observed distinctive site-specific changes in glycosylation. Compared to previous studies on the human milk glycoproteome, we took a more integral approach by focusing on intact glycopeptides and were able to reach considerably more depth. Other studies either focused on the released *N*-glycans^{15, 16} or deglycosylated peptides¹²⁻¹⁴. We consider the improvement in depth in our study to originate mainly from two factors: 1) the application of the robust and reproducible automated HILIC-based enrichment protocol, and 2) the high-end high-resolution mass spectrometry approach using HCD-triggered EThcD-based glycopeptide identification. The automated HILIC-based enrichment on the AssayMAP Bravo robot has the capacity to allow for the enrichment of 96 digested samples in parallel within 1 hour, offering the potential for application on studies with numerous samples, such as the longitudinal time-series presented here. We demonstrate that it is now possible to monitor personalized milk glycoproteomes, i.e. per individual donor. By doing this, our data revealed a fascinating variability in glycans and their site-occupancy, whereby each human milk glycoprotein, each *N*-glycosylation site and each glycan attached exhibit their own distinctive quantitative pattern across the monitored lactation period. All these observed variations are very likely related to the changing needs of the growing and developing infant, and are hallmarked by the changing functional demands of the immune system and diversifying gut microbiome.

Supporting Information

The following supporting information is available free of charge at ACS website <http://pubs.acs.org>

Supplementary FIG. 1 Dynamic range of human milk proteome and identified glycoproteins.

Supplementary FIG. 2 Global comparison of the glycan composition of *N*-glycopeptides in human milk vs human serum

Supplementary FIG. 3 Represented site occupancy of Asn69 from alpha-S1-casein by one major glycopeptide and unmodified peptide.

Supplementary FIG. 4 Protein sequences alignment of alpha-S1-casein across different species.

Supplementary FIG. 5 Experimental parameters used in the scheduled targeted assay for glycopeptide quantification.

Supplementary File: Visual summary for each identified glycoprotein. Schematic drawing of the sequences of the identified glycoproteins with identified sites and identified different glycan compositions.

Supplementary Table S1: 200 possible N-glycan compositions in human without sodium and with multiple fucoses

Supplementary Table S2: Peptide to spectrum matches for EThcD spectra

Supplementary Table S3: Identified N-glycopeptides in human milk across lactation

Supplementary Table S4: Glycan compositions at identified N-glycosites

Supplementary Table S5: The identities of the outer nodes in the glycoprotein-glycan network diagram (Figure 2C)

Supplementary Table S6: The normalized intensity of glycopeptides across lactation in one single donor

Acknowledgments

We acknowledge support from the Netherlands Organization for Scientific Research (NWO) funding the large-scale proteomics facility *Proteins@Work* (project 184.032.201) embedded in the Netherlands Proteomics Centre. A.J.R.H. acknowledge further support by the NWO TOP-Punt Grant 718.015.003 and the and EU Horizon 2020 program INFRAIA project Epic-XS (Project 823839). J.Z. acknowledges support from the Chinese Scholarship Council (CSC). Additional support for this research was provided by Danone Nutricia Research. We thank Vojtech Franc and Karli Reiding (Utrecht University) for proofreading the manuscript.

Conflicts of Interest

M.M and B.S are employees of Danone Nutricia Research. J.Z and K.A.D were enrolled as PhD students at Utrecht University during this study and received partial financial support from Danone Nutricia Research. None of the authors have further conflicts of interest with regard to the content of this manuscript.

Data Availability

The mass spectrometry data have been deposited to the ProteomeXchange Consortium via the PRIDE⁴³ partner repository with the dataset identifier PXD013764 and 10.6019/PXD013764.

Skyline data have been deposited to the ProteomeXchange Consortium via Panorama⁴⁴ with the access URL (<https://panoramaweb.org/BBgb9c.url>).

References

1. Hennet, T.; Borsig, L., Breastfed at Tiffany's. *Trends Biochem Sci* 2016, 41 (6), 508-518.
2. Newburg, D. S., Glycobiology of human milk. *Biochemistry (Mosc)* 2013, 78 (7), 771-85.
3. Lis-Kuberka, J.; Orczyk-Pawilowicz, M., Sialylated Oligosaccharides and Glycoconjugates of Human Milk. The Impact on Infant and Newborn Protection, Development and Well-Being. *Nutrients* 2019, 11 (2), 306.
4. Hayashi, H.; Yamashita, Y., Role of N-glycosylation in cell surface expression and protection against proteolysis of the intestinal anion exchanger SLC26A3. *Am J Physiol Cell Physiol* 2012, 302 (5), C781-95.
5. Liu, B.; Newburg, D. S., Human milk glycoproteins protect infants against human pathogens. *Breastfeed Med* 2013, 8 (4), 354-62.
6. Dixon, D. L., The Role of Human Milk Immunomodulators in Protecting Against Viral Bronchiolitis and Development of Chronic Wheezing Illness. *Children (Basel)* 2015, 2 (3), 289-304.
7. Georgi, G.; Bartke, N.; Wiens, F.; Stahl, B., Functional glycans and glycoconjugates in human milk. *Am J Clin Nutr* 2013, 98 (2), 578S-85S.
8. Fields, D. A.; Schneider, C. R.; Pavela, G., A narrative review of the associations between six bioactive components in breast milk and infant adiposity. *Obesity (Silver Spring)* 2016, 24 (6), 1213-21.
9. Zhu, J.; Garrigues, L.; Van den Toorn, H.; Stahl, B.; Heck, A. J. R., Discovery and quantification of nonhuman proteins in human milk. *Journal of proteome research* 2018, 18 (1), 225-238.
10. Caval, T.; Tian, W.; Yang, Z.; Clausen, H.; Heck, A. J. R., Direct quality control of glycoengineered erythropoietin variants. *Nat Commun* 2018, 9 (1), 3342.
11. Caval, T.; Zhu, J.; Heck, A. J. R., Simply Extending the Mass Range in Electron Transfer Higher Energy Collisional Dissociation Increases Confidence in N-Glycopeptide Identification. *Analytical chemistry* 2019, 91 (16), 10401-10406.

12. Cao, X.; Yang, M.; Yang, N.; Liang, X.; Tao, D.; Liu, B.; Wu, J.; Yue, X., Characterization and comparison of whey N-glycoproteomes from human and bovine colostrum and mature milk. *Food Chem* 2019, 276, 266-273.
13. Cao, X.; Kang, S.; Yang, M.; Li, W.; Wu, S.; Han, H.; Meng, L.; Wu, R.; Yue, X., Quantitative N-glycoproteomics of milk fat globule membrane in human colostrum and mature milk reveals changes in protein glycosylation during lactation. *Food Funct* 2018, 9 (2), 1163-1172.
14. Picariello, G.; Ferranti, P.; Mamone, G.; Roepstorff, P.; Addeo, F., Identification of N-linked glycoproteins in human milk by hydrophilic interaction liquid chromatography and mass spectrometry. *Proteomics* 2008, 8 (18), 3833-47.
15. Nwosu, C. C.; Aldredge, D. L.; Lee, H.; Lerno, L. A.; Zivkovic, A. M.; German, J. B.; Lebrilla, C. B., Comparison of the human and bovine milk N-glycome via high-performance microfluidic chip liquid chromatography and tandem mass spectrometry. *Journal of proteome research* 2012, 11 (5), 2912-24.
16. Dallas, D. C.; Martin, W. F.; Strum, J. S.; Zivkovic, A. M.; Smilowitz, J. T.; Underwood, M. A.; Affolter, M.; Lebrilla, C. B.; German, J. B., N-linked glycan profiling of mature human milk by high-performance microfluidic chip liquid chromatography time-of-flight tandem mass spectrometry. *J Agric Food Chem* 2011, 59 (8), 4255-63.
17. Yoshida, Y.; Furukawa, J. I.; Naito, S.; Higashino, K.; Numata, Y.; Shinohara, Y., Quantitative analysis of total serum glycome in human and mouse. *Proteomics* 2016, 16 (21), 2747-2758.
18. Bekker-Jensen, D. B.; Kelstrup, C. D.; Bath, T. S.; Larsen, S. C.; Haldrup, C.; Bramsen, J. B.; Sorensen, K. D.; Hoyer, S.; Orntoft, T. F.; Andersen, C. L.; Nielsen, M. L.; Olsen, J. V., An Optimized Shotgun Strategy for the Rapid Generation of Comprehensive Human Proteomes. *Cell Syst* 2017, 4 (6), 587-599 e4.
19. MacLean, B.; Tomazela, D. M.; Shulman, N.; Chambers, M.; Finney, G. L.; Frewen, B.; Kern, R.; Tabb, D. L.; Liebler, D. C.; MacCoss, M. J., Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 2010, 26 (7), 966-8.
20. Lee, L. Y.; Moh, E. S.; Parker, B. L.; Bern, M.; Packer, N. H.; Thaysen-Andersen, M., Toward Automated N-Glycopeptide Identification in Glycoproteomics. *Journal of proteome research* 2016, 15 (10), 3904-3915.

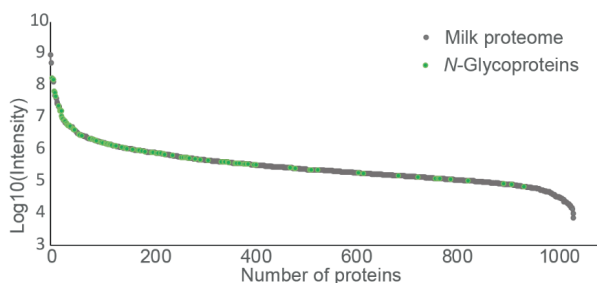
21. Riley, N. M.; Hebert, A. S.; Westphall, M. S.; Coon, J. J., Capturing site-specific heterogeneity with large-scale N-glycoproteome analysis. *Nat Commun* 2019, 10 (1), 1311.
22. Bai, Y.; Tao, J.; Zhou, J.; Fan, Q.; Liu, M.; Hu, Y.; Xu, Y.; Zhang, L.; Yuan, J.; Li, W.; Ze, X.; Malard, P.; Guo, Z.; Yan, J.; Li, M., Fucosylated Human Milk Oligosaccharides and N-Glycans in the Milk of Chinese Mothers Regulate the Gut Microbiome of Their Breast-Fed Infants during Different Lactation Stages. *mSystems* 2018, 3 (6), e00206-18.
23. Sun, S.; Hu, Y.; Jia, L.; Eshghi, S. T.; Liu, Y.; Shah, P.; Zhang, H., Site-Specific Profiling of Serum Glycoproteins Using N-Linked Glycan and Glycosite Analysis Revealing Atypical N-Glycosylation Sites on Albumin and alpha-1B-Glycoprotein. *Analytical chemistry* 2018, 90 (10), 6292-6299.
24. Fouda, G. G.; Jaeger, F. H.; Amos, J. D.; Ho, C.; Kunz, E. L.; Anasti, K.; Stamper, L. W.; Liebl, B. E.; Barbas, K. H.; Ohashi, T., Tenascin-C is an innate broad-spectrum, HIV-1-neutralizing protein in breast milk. *Proceedings of the National Academy of Sciences* 2013, 110 (45), 18220-18225.
25. Liao, Y.; Weber, D.; Xu, W.; Durbin-Johnson, B. P.; Phinney, B. S.; Lonnerdal, B., Absolute Quantification of Human Milk Caseins and the Whey/Casein Ratio during the First Year of Lactation. *Journal of proteome research* 2017, 16 (11), 4113-4121.
26. Konig, S.; Altendorfer, I.; Saenger, T.; Bleck, E.; Vordenbaumen, S.; Schneider, M.; Jose, J., Ser71 of alphaS1-Casein is Phosphorylated in Breast Milk-Evidence from Targeted Mass Analysis. *Molecular nutrition & food research* 2017, 61 (12).
27. Johnsen, L. B.; Rasmussen, L. K.; Petersen, T. E.; Berglund, L., Characterization of three types of human alpha s1-casein mRNA transcripts. *The Biochemical journal* 1995, 309 (Pt 1), 237-42.
28. Zeng, W.-F.; Liu, M.-Q.; Zhang, Y.; Wu, J.-Q.; Fang, P.; Peng, C.; Nie, A.; Yan, G.; Cao, W.; Liu, C.; Chi, H.; Sun, R.-X.; Wong, C. C. L.; He, S.-M.; Yang, P., pGlyco: a pipeline for the identification of intact N-glycopeptides by using HCD- and CID-MS/MS and MS3. *Scientific Reports* 2016, 6, 25102.
29. Elwakiel, M.; Boeren, S.; Hageman, J. A.; Szeto, I. M.; Schols, H. A.; Hettinga, K. A., Variability of Serum Proteins in Chinese and Dutch Human Milk during Lactation. *Nutrients* 2019, 11 (3), 499.
30. Palmeira, P.; Carneiro-Sampaio, M., Immunology of breast milk. *Rev Assoc Med Bras (1992)* 2016, 62 (6), 584-593.

31. Phalipon, A.; Cortesy, B., Novel functions of the polymeric Ig receptor: well beyond transport of immunoglobulins. *Trends Immunol* 2003, 24 (2), 55-8.
32. Landberg, E.; Huang, Y.; Stromqvist, M.; Mechref, Y.; Hansson, L.; Lundblad, A.; Novotny, M. V.; Pahlsson, P., Changes in glycosylation of human bile-salt-stimulated lipase during lactation. *Arch Biochem Biophys* 2000, 377 (2), 246-54.
33. Barboza, M.; Pinzon, J.; Wickramasinghe, S.; Froehlich, J. W.; Moeller, I.; Smilowitz, J. T.; Ruhaak, L. R.; Huang, J.; Lonnerdal, B.; German, J. B.; Medrano, J. F.; Weimer, B. C.; Lebrilla, C. B., Glycosylation of human milk lactoferrin exhibits dynamic changes during early lactation enhancing its role in pathogenic bacteria-host interactions. *Mol Cell Proteomics* 2012, 11 (6), M111 015248.
34. Schneider, M.; Al-Shareffi, E.; Haltiwanger, R. S., Biological functions of fucose in mammals. *Glycobiology* 2017, 27 (7), 601-618.
35. Doherty, A. M.; Lodge, C. J.; Dharmage, S. C.; Dai, X.; Bode, L.; Lowe, A. J., Human Milk Oligosaccharides and Associations With Immune-Mediated Disease and Infection in Childhood: A Systematic Review. *Front Pediatr* 2018, 6, 91.
36. Donovan, S. M.; Comstock, S. S., Human Milk Oligosaccharides Influence Neonatal Mucosal and Systemic Immunity. *Ann Nutr Metab* 2016, 69, 42-51.
37. Mezzoff, E. A.; Hawkins, J. A.; Ollberding, N. J.; Karns, R.; Morrow, A. L.; Helmrath, M. A., The human milk oligosaccharide 2'-fucosyllactose augments the adaptive response to extensive intestinal. *Am J Physiol-Gastr L* 2016, 310 (6), G427-G438.
38. Garrido, D.; Nwosu, C.; Ruiz-Moyano, S.; Aldredge, D.; German, J. B.; Lebrilla, C. B.; Mills, D. A., Endo-beta-N-acetylglucosaminidases from Infant Gut-associated Bifidobacteria Release Complex N-glycans from Human Milk Glycoproteins. *Molecular & Cellular Proteomics* 2012, 11 (9), 775-785.
39. Li, M.; Bai, Y.; Zhou, J.; Huang, W.; Yan, J.; Tao, J.; Fan, Q.; Liu, Y.; Mei, D.; Yan, Q.; Yuan, J.; Malard, P.; Wang, Z.; Gu, J.; Taniguchi, N.; Li, W., Core Fucosylation of Maternal Milk N-Glycan Evokes B Cell Activation by Selectively Promoting the 1-Fucose Metabolism of Gut Bifidobacterium spp. and Lactobacillus spp. *MBio* 2019, 10 (2), e00128-19.
40. Karav, S.; Le Parc, A.; Bell, J. M. L. N. D.; Frese, S. A.; Kirmiz, N.; Block, D. E.; Barile, D.; Mills, D. A., Oligosaccharides Released from Milk Glycoproteins Are Selective Growth Substrates for Infant-Associated Bifidobacteria. *Appl Environ Microb* 2016, 82 (12), 3622-3630.

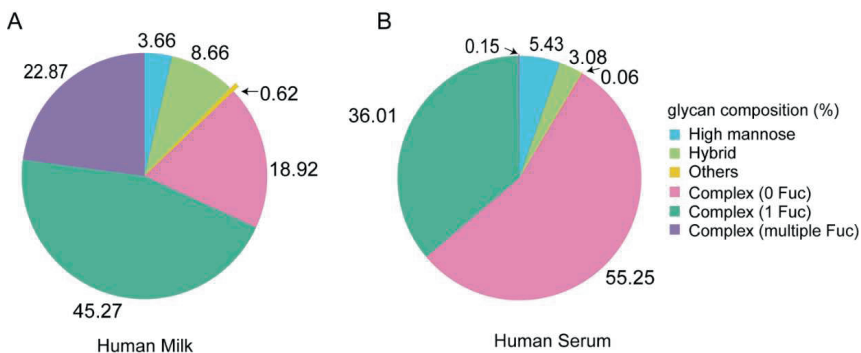
41. Lonnerdal, B., Bioactive Proteins in Human Milk: Health, Nutrition, and Implications for Infant Formulas. *J Pediatr-Us* 2016, 173, S4-S9.
42. Maruyama, K.; Hida, M.; Kohgo, T.; Fukunaga, Y., Changes in salivary and fecal secretory IgA in infants under different feeding regimens. *Pediatr Int* 2009, 51 (3), 342-5.
43. Vizcaino, J. A.; Csordas, A.; Del-Toro, N.; Dianes, J. A.; Griss, J.; Lavidas, I.; Mayer, G.; Perez-Riverol, Y.; Reisinger, F.; Ternent, T.; Xu, Q. W.; Wang, R.; Hermjakob, H., 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res* 2016, 44 (22), 11033.
44. Sharma, V.; Eckels, J.; Taylor, G. K.; Shulman, N. J.; Stergachis, A. B.; Joyner, S. A.; Yan, P.; Whiteaker, J. R.; Halusa, G. N.; Schilling, B.; Gibson, B. W.; Colangelo, C. M.; Paulovich, A. G.; Carr, S. A.; Jaffe, J. D.; MacCoss, M. J.; MacLean, B., Panorama: a targeted proteomics knowledge base. *Journal of proteome research* 2014, 13 (9), 4205-10.

Supporting information for Chapter 5

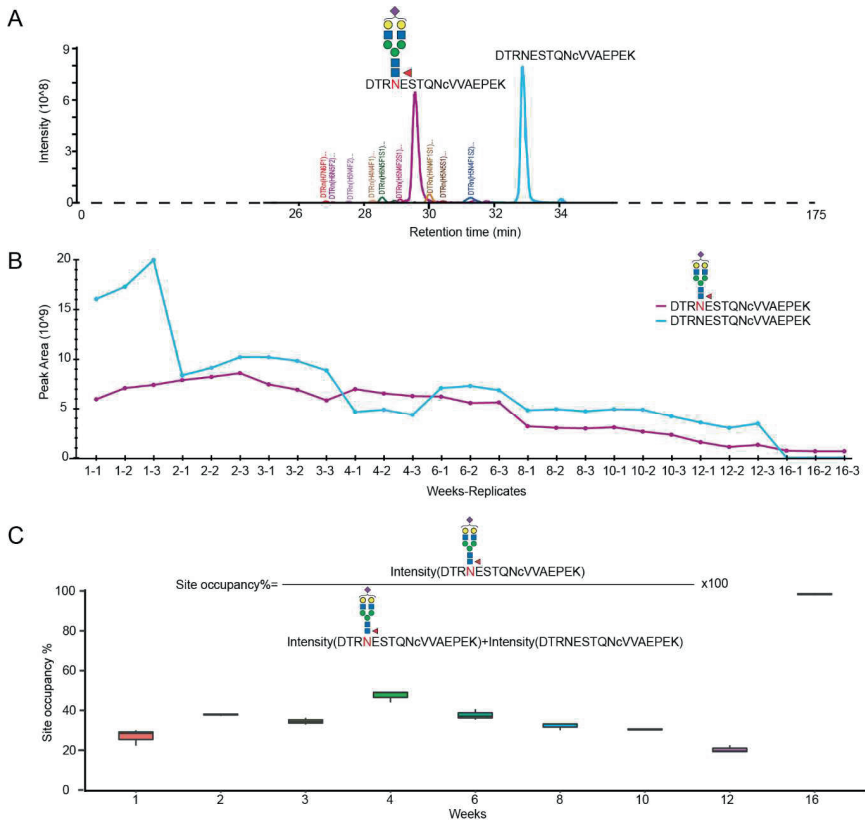
Supplementary Figures



Supplementary FIG. 1 Dynamic range in the human milk proteome and identified glycoproteins. The dynamic range of detected glycoproteins spans the whole range of proteins detected in the human milk proteome. Proteins detected in the human milk proteome are plotted by their intensity in grey and the glycoproteins identified are mapped on top of them and colored in green.



Supplementary FIG. 2 Global comparison of the glycan composition of N-glycopeptides in human milk vs human serum. A. Composition of N-linked glycoforms by PSMs in human milk based on PSMs (this study). B. Composition of N-linked glycoforms by PSMs in human serum based on data from Sun *et al.*²¹. In particular, hybrid structures and complex glycans harboring multiple fucoses seems to be more abundant in human milk vs serum.



Supplementary FIG. 3 Represented N-glycan site occupancy of Asn69 on alpha-S1-casein as extracted from data from one major abundant glycopeptide and the unmodified peptide. A. An illustrative example of XICs of the unmodified peptide DTRNESTQNCVVAEPEK and its glycosylated form harboring a HexNAc(4)Hex(5)Fuc(1)Sia(1) glycan and other less abundant glycosylated forms. B. The XICs of the unmodified peptide DTRNESTQNCVVAEPEK and its glycosylated form harboring a HexNAc(4)Hex(5)Fuc(1)Sia(1) glycan in all replicates. C. Observed site occupancy across the stages of lactation measured in triplicates. The glycosylation site occupancies ranged from 20% (weeks 1 and 12) to almost fully occupied 95% (week 16).

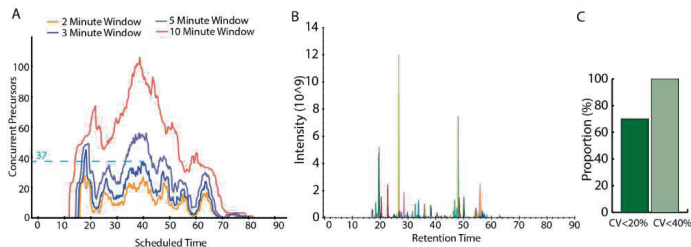
```

SP|P47710-1|CASA1_HUMAN  MRLLLITCLVAVALARPKLPLRYPERLQNPSE-----SSE-PIPLESR 42
SP|P47710-2|CASA1_HUMAN  MRLLLITCLVAVALARPKLPLRYPERLQNPSE-----SSE-PIPLESR 42
SP|P47710-3|CASA1_HUMAN  MRLLLITCLVAVALARPKLPLRYPERLQNPSE-----SSE-PIPLESR 42
SP|P47710-4|CASA1_HUMAN  MRLLLITCLVAVALARPKLPLRYPERLQNPSE-----SSE-PIPLESR 42
SP|P02662|CASA1_BOVIN    MKLLIITCLVAVALARPKHPKIKHQGLPQE-----VLN-ENLLRFVAPFPEVFG 48
SP|P19228|CASA1_MOUSE    MKLLIITCLVAAAFAMPRLHSRNVSSQTQQQHSSS--EEIFKQ-----PKYLNLN 49
SP|P04653|CASA1_SHEEP    SP|P04653|CASA1_SHEEP    MKLLIITCLVAVALARPKHPKIKHQGLSSE-----VLN-ENLLRFVAPFPEVFR 48
SP|P02661|CASA1_RAT      MKLLIITCLVAAALALPRAHRRNVSSQTQQENSSSEEQEIVKQ-----PKYLSLN 51
SP|P18626|CASA1_CAPHI    MKLLIITCLVAVALARPKHPINHRLSPSE-----VFN-ENLLRFVAPFPEVFR 48
SP|O97943|CASA1_CAMDR    MKLLIITCLVAVALARPKYPLRYPEVFQNEPDSIE---EVLNKRKILELAVVSP-IQFR 55
SP|P39035|CASA1_PIG      MKLLIFICLAAVALARPKPPLRHQEHQNEPDSRE---ELFKERKFLRFPEVPLLSQFR 56
SP|P86272|CASA1_EQUAS    -----RPKLPHRHPEIIQNEQDSRE-----KVLKERKFPFALHTPRE--- 38
SP|P09115|CASA1_RABIT    MKLLIITCLVATALARHFKFHLGHLKLTQEQPESSE--QEILKERKLLRFVQ--TVPLELR 56
SP|P28549|CASA1_MACEU    MKLLIFSLCVTLALARPDALRLSIDRHFHKREL---ENRLNEDPI-----PVSEASSS 50
SP|P04656|CASA1_CAVPO    MKLLIITCLVASAVAMPKFPFRHTELFQTRQGSSS-----SSSSSEE-----RLK 45
SP|O62823|CASA1_BUBBU    MKLLIITCLVAVALARPKQPIKHQGLPQG-----VLN-ENLLRFVAPFPEVFG 48

SP|P47710-1|CASA1_HUMAN  EEYMGMRNRQNRNLLREKQTDKIDTRNEST-QNCVVAEPEKM-----ESSISS- 90
SP|P47710-2|CASA1_HUMAN  EEYMGMRNR-RNLLREKQTDKIDTRNEST-QNCVVAEPEKM-----ESSISS- 89
SP|P47710-3|CASA1_HUMAN  EEYMGMRNRQNRNLLREKQTDKIDTRNEST-QNCVVAEPEKM-----ESSISS- 82
SP|P47710-4|CASA1_HUMAN  EEYMGMRNR-RNLLREKQTDKIDTRNEST-QNCVVAEPEKM-----ESSISS- 89
SP|P02662|CASA1_BOVIN    KEKVNELSKD-----IGSES-----T-EDQAMEDIKQM-----EASISS- 82
SP|P19228|CASA1_MOUSE    QEFVNMMNRQRALLTEQ-NDEIKVTMDAAS-EEQAMASAQED-----SSIS-SSS 96
SP|P04653|CASA1_SHEEP    SP|P04653|CASA1_SHEEP    KENINELSKD-----IGSES-----I-EDQAMEDIKQM-----KAGSSSS- 82
SP|P02661|CASA1_RAT      EEFVNNLNRQRELLTEQ-DNEIKITMDSSA-EEQATASAQED-----SSSSSSSS 99
SP|P18626|CASA1_CAPHI    KENINELSKD-----IGSES-----T-EDQAMEDIKQM-----KAGSSSS- 82
SP|O97943|CASA1_CAMDR    QENIDELKDRN-----EP-----T-EDHIMEDTERK-----E-SGSS- 87
SP|P39035|CASA1_PIG      QEILNELN-----RNHGMGHEQR-----G-SSSS- 81
SP|P86272|CASA1_EQUAS    -EYINELNRQRELLKEKQKDE-----H-KEYLIEDPEQQ-----ESSSTSS- 77
SP|P09115|CASA1_RABIT    EEVYNELNRQRELLREKENEEIKGTRNEVT-EEHVLADRET-----EASISS- 103
SP|P28549|CASA1_MACEU    EESVHQLNRDRRPLEKYELDLYREDLKTSSSEEFVTPSTNERNVRRQVEYFNFEEDSSAS- 109
SP|P04656|CASA1_CAVPO    EENIFKFPDQKEL-QRKQSEKIKEIIS-----ESTFQR-----EASISS- 84
SP|O62823|CASA1_BUBBU    KEKVNELSTD-----IGSES-----T-EDQAMEDIKQM-----EASISS- 82

```

Supplementary FIG. 4 Protein sequence alignment for alpha-S1-casein across different species. The red colored letters indicate the novel *N*-glycosylation site found in human alpha-S1-casein. This site is not present in isoform 3 of human alpha-S1-casein. It seems to be unique for the human orthologue and is not present in related mammalian species.



Supplementary FIG. 5 Experimental parameters used in the scheduled targeted assay for glycopeptide quantification. A. With the chosen 3 minutes window, the concurrent precursors were at a maximum 37 (blue line). Taking 54 ms as the ion injection time, a mass resolution setting of 30,000 and an average peak width of 30 s, at least 10 data points were taken over each chromatographic peak. B. Overall chromatogram of targeted glycopeptides. C. The CVs of all spiked PRTC peptides were calculated for all the runs, 70% of them had a CV less than 20%, another 30% had a CV less than than 40%. CV: coefficient of variation; PRTC: Pierce Peptide Retention Time Calibration Mixture.

The following supplementary tables are not printed because they are oversize.

Supplementary File: Visual summary for each identified glycoprotein.

Schematic drawing of the sequences of the identified glycoproteins with identified sites and identified different glycan compositions. In total, 110 proteins are depicted in this file, of which 51 showed altered changes of certain glycan composition over the 9 measured time points. The protein sequence chains are color-coded, wherein light blue, blue and grey represent the signal peptide, mature protein and propeptide chains, respectively. Indigo dots inserted in the protein sequences with and without asterisk indicate the UniProt-annotated N-glycosylation sites, either reported as being with evidence or only based on sequon analysis, respectively. Purple dots indicate the novel reported glycosites in this study. A red circle indicates that the glycosite is identified in this study. Yellow dots with a red circle highlight those glycosites reported in the UniProt database only based on sequon analysis, but here confirmed by experimental evidence in our dataset.

Supplementary Table S1: 200 possible N-glycan compositions in human without sodium and with multiple fucoses

Supplementary Table S2: Peptide to spectrum matches for EThcD spectra

Supplementary Table S3: Identified N-glycopeptides in human milk across lactation

Supplementary Table S4: Glycan compositions at identified N-glycosites

Supplementary Table S5: The identities of the outer nodes in the glycoprotein-glycan network diagram (Figure 2C)

Supplementary Table S6: The normalized intensity of glycopeptides across lactation in one single donor

Chapter 6

Neutrophil azurophilic granule proteins are distinctively decorated by highly atypical pauci- and phosphomannose glycans

Karli R. Reiding^{ah,sk*}, Yu-Hsien Lin^{ah,s}, Albert J. R. Heck^{ah,*}

[†]Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Padualaan 8, 3584 CH, Utrecht, The Netherlands;

[§]Netherlands Proteomics Center, Padualaan 8, 3584 CH, Utrecht, The Netherlands;

[¶]These authors contributed equally

*Correspondence may be addressed to Karli R. Reiding (k.r.reiding@uu.nl) or Albert J.R. Heck (a.j.r.heck@uu.nl)

Short title: Site-specific neutrophil *N*-glycoproteomics

Manuscript submitted

Abstract

Neutrophils are the critical first-responders of the immune system, but may also cause off-target damage and appear involved in a variety of autoimmune diseases. Many neutrophil proteins are *N*-glycosylated, a post-translational modification that significantly affects, among others, enzymatic activity, receptor interaction, and backbone accessibility. Interestingly, several neutrophil proteins were found to carry unusually small glycans (paucimannose) and those with phosphorylation, and this could differ highly from one glycosylation site to the next. To uncover the scope of these uncommon glycoforms across the neutrophil proteome, we performed in-depth LC-MS/MS-based (glyco)proteomics of the combined neutrophils from healthy donors. Achieving shotgun detection of 5444 proteins and site-specific *N*-glycan characterization of 241, we found that membrane proteins expressed high-mannose or complex *N*-glycans depending on the site, and secreted proteins mostly just complex *N*-glycans. Notably, on proteins likely to originate from azurophilic granules, we report distinct and dominant paucimannosylation, asymmetry, and glycan phosphorylation.

Keywords

Neutrophil, proteome, proteomics, glycoproteome, glycoproteomics, glycosylation, paucimannose, phosphomannose, mass spectrometry

Introduction

Neutrophils play a critical role in the immune system as they are one of the first responders to migrate towards sites of inflammation and perform local antimicrobial functions^{1,2}. Underlining their importance, neutrophils are produced in massive quantities and end up filling more than half the population of leukocytes in human blood – even though the individual cells are relatively short-lived and can be measured in the order of days³. Neutrophils perform their action *via* a number of mechanisms, including the phagocytosis of pathogens and subsequent digestion in intracellular phagosomes, the ensnaring of pathogens by releasing DNA in the form of neutrophil extracellular traps (NETs), or by the release of toxic proteins mixtures into the extracellular space in a process called degranulation⁴. This degranulation can occur from several distinct granules, including azurophilic (or primary) granules, specific (or secondary) granules and gelatinase (or tertiary) granules, each loaded with a particular set of proteins^{5,6}. Azurophilic granules, for instance, are known to contain the abundant peroxidase myeloperoxidase as well as the proteases neutrophil elastase and myeloblastin (proteinase 3), while the iron-scavenger lactotransferrin and glycosidase lysozyme C are primarily contained within specific granules^{5,6}.

While neutrophils are essential to protect against invading pathogens, the cells are unfortunately not very discriminatory once their defenses have been activated. Next to handling pathogens, the degranulate inadvertently destroy otherwise healthy proteins and carbohydrates⁷, leading to lasting damage to the extracellular matrix and tissue and contributing to disease phenotypes such as asthma and chronic obstructive pulmonary disease (COPD)⁸. Notably, a subset of neutrophil proteins also tends to develop into the target of an autoimmune response, initiated by the collectively-named anti-neutrophil cytoplasmic autoantibodies (ANCA)s⁹⁻¹¹. These ANCA are found in multiple autoimmune diseases, including rheumatoid arthritis and inflammatory bowel disease, but they are most notable for their diagnostic value in distinguishing subtypes of ANCA-associated vasculitides (AAV)⁹⁻¹¹. Even while neutrophils are of obvious clinical interest in terms treating disease complications and uncovering epitopes of autoantigenicity, very little is actually known about the post-translational modifications (PTMs) that naturally occur on the cell's proteome.

One frequent PTM, glycosylation, represents the directed modification of proteins with one or more complex sugars, which can then be remodeled into a large degree of functionally-relevant variation¹². Glycosylation has a major effect on a protein as the sugar affects, among other things, the folding and receptor interaction of the protein, as well as the availability of the peptide backbone to proteases and the recognition by autoantibodies¹²⁻¹⁵. The glycosylation characteristics of neutrophils are not well-known, save for reports on neutrophil elastase, cathepsin G, and myeloblastin¹⁶⁻¹⁸, while an early *N*-glycoproteomics experiment on

(neutrophil-rich) sputum detected 115 *N*-glycopeptides that likely originated from neutrophils¹⁹. Interestingly, the overall indication is that a subset of neutrophil proteins likely contains an infrequently observed class of small carbohydrates – paucimannose glycans^{19,20}. Adding to this, our group recently performed a thorough mass spectrometric characterization of the abundant neutrophil protein myeloperoxidase²¹, uncovering that the protein expressed, indeed next to the aforementioned paucimannose glycans, a significant amount of phosphomannosylation (Asn323) and highly asymmetric species (Asn483), and that this expression differed significantly depending on the glycosylation site²¹. The question remains if these uncommon glycosylation characteristics are typical for neutrophil proteins in general, or represent a relevant subset thereof.

Here, by means of the latest innovations in sample preparation, hybrid mass spectrometry (MS) and data analysis²²⁻²⁴, we present the first in-depth site-specific characterization of the neutrophil *N*-glycoproteome. Starting from a pool of neutrophils mixed from ten independent donors, we achieved the MS-based detection and label-free quantification of 5444 neutrophil proteins, and notably, site-specific *N*-glycosylation information on 241. By this, we found that membrane proteins expressed either high-mannose glycans or complex glycans with high levels of fucosylation, likely Lewis-X (CD15) and sialyl-Lewis X (CD15s), while secretory proteins mostly displayed complex glycans only. Importantly, we indeed found abundant paucimannose, phosphomannose and asymmetrical glycan species, but primarily amongst those proteins that typically reside in the azuophilic granules. We believe that this report will aid in understanding neutrophil biology and the pathogenicity that may arise from it, as well as informing on leads for targeted intervention within neutrophil-based malignancies.

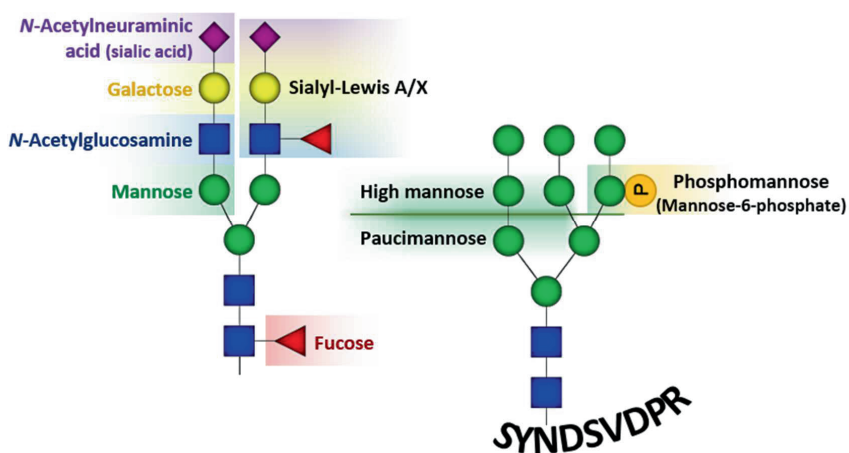


Figure 1: Annotations used to describe glycan compositions. Monosaccharides are represented as: fucose (Fuc; red triangle), mannose (Man; green circle), galactose (Gal; yellow circle), *N*-acetylglucosamine (GlcNAc; blue square) and *N*-acetylneuraminic acid (NeuAc; dark magenta diamond), phosphomannose (Phospho; orange circle). For more complex glycans (left) distinctions are made between asymmetric/hybrid (one antennary GlcNAc), diantennary (two antennary GlcNAcs) and extended species (more than two antennary GlcNAcs). For mannose-type glycans (right) classes are defined as paucimannose (three mannoses or less), high-mannose (four mannoses or more), and phosphomannosylated glycans (carrying at least one phosphomannose).

Results

To uncover the site-specific *N*-glycosylation characteristics of neutrophil proteins (**Figure 1**), we pooled the neutrophils from ten presumably healthy donors to perform in-depth proteomics and glycoproteomics analyses. Cells were lysed and the proteins thereof reduced, alkylated and proteolytically digested by GluC and trypsin.

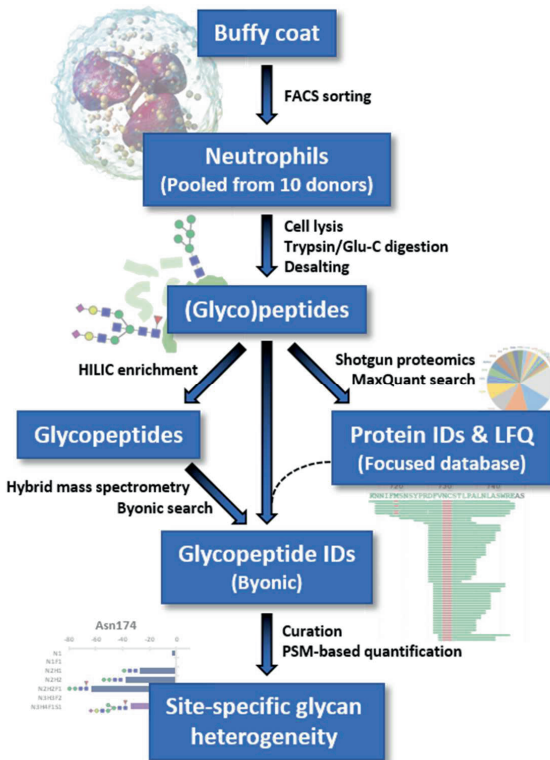


Figure 2: Scheme of the experimental workflow. Neutrophils, acquired by flow cytometry of buffy coats, were pooled from 10 independent, presumably healthy, donors. The cells were disrupted, the released proteins digested, and the peptides desalted by reversed phase (RP) solid phase enrichment (SPE). The resulting (glyco)peptides were then either further enriched by hydrophilic-interaction liquid chromatography (HILIC) SPE before mass spectrometry (MS) analysis (**left branch**), or directly, without enrichment, analyzed by MS (**right branch**).

Results from shotgun proteomics (desalting-only) were used for the identification and quantification (LFQ) of the proteins, whereas the glycoproteomics data resulting from both desalting-only and desalting+HILIC were used for identification and quantification of the glycopeptides.

One part of the resulting peptides was desalted by reversed phase (RP)-solid phase enrichment (SPE) only, whereas another part was, next to desalting, also enriched for glycopeptides by hydrophilic-interaction liquid chromatography (HILIC)-SPE (**Figure 2**).

Shotgun proteomics

We performed shotgun proteomics on the RP-enriched neutrophil (glycopeptides) by RP-LC-MS². The resulting MS data was searched against the reviewed human proteome (Swiss-Prot database) allowing either 1) specific peptides up to 3 miscleavages, or 2) complete nonspecificity.

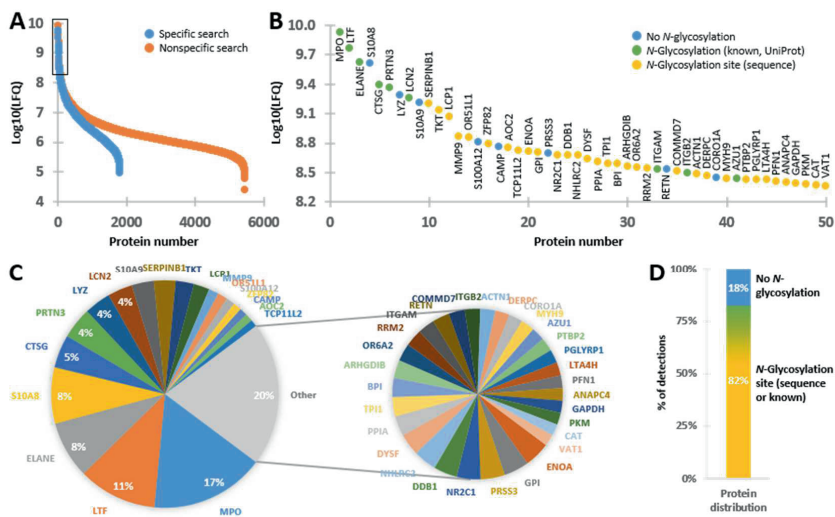


Figure 3: Quantitative human neutrophil proteome. **A)** Relative quantification of the 1802 neutrophil proteins detected with peptides identified from a specific trypsin/GluC search (blue). Allowing also nonspecific cleavages in the search led to the detection of 5444 proteins (orange). **B)** Close-up of the top 50 most abundant proteins. The colours indicate whether the known protein sequence contained either no *N*-glycosylation site (blue; sites being defined as Asn-Xxx-Ser/Thr where Xxx ≠ Pro), at least one site with evidence for *N*-glycosylation as annotated in UniProt (green), or at least one site predicted by sequence analysis (yellow). **C)** Relative abundance of the top 50 most abundant proteins. **D)** *N*-Glycosylation site distribution across all detected proteins (nonspecific).

The specific search yielded the detection and label-free quantification (LFQ) of 1802 proteins (Figure 3), abundant examples including the common neutrophil proteins lactotransferrin, myeloperoxidase, as well as protein S100-A8 and A9 which typically heterooligomerise into calprotectin (Supplementary Table S1)²⁵. These proteins were in congruence with those from other neutrophil proteomics studies, and indicated a broad detection of proteins from different cellular compartments^{5,25,26}. From the literature, we could infer that our detections included proteins from the membrane, cytosol, azurophilic (primary) granules, specific (secondary) granules, gelatinase (tertiary) granules, and secretory vesicles⁵.

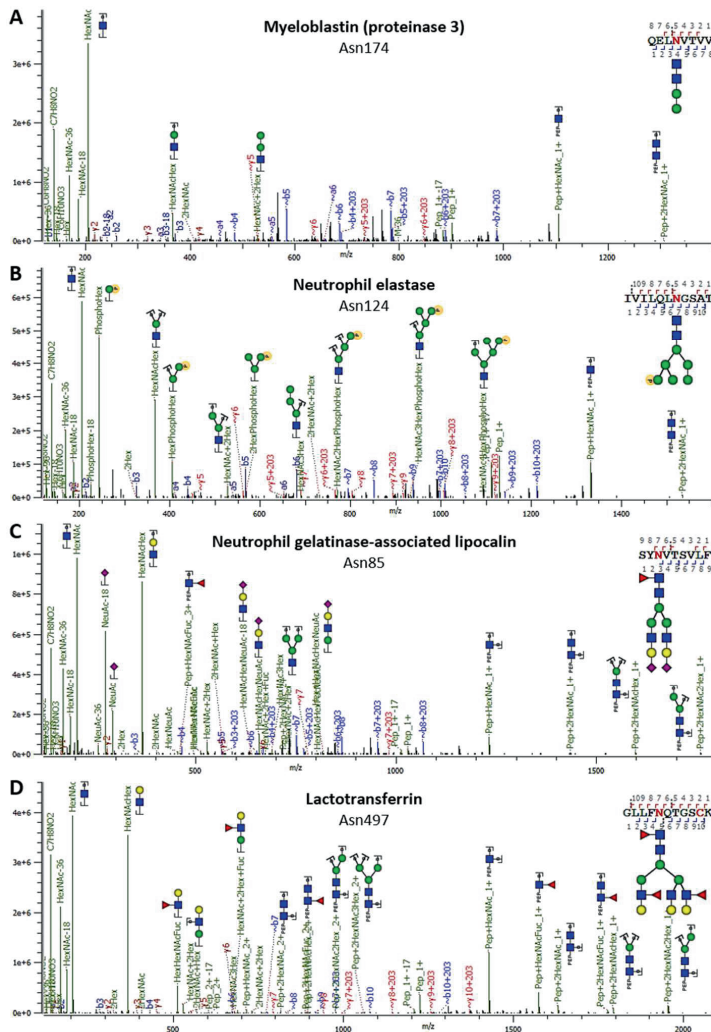


Figure 4: Prototypical MS/MS spectra exemplifying the different categories of glycosylation observed in neutrophils. A) paucimannosylation, B) phosphomannosylation, C)

sialylation, and **D**) high degrees of (antennary) fucosylation. In **A**) the fragments b4+203 and y5+203 position the paucimannosylated glycan unambiguously at the Asn residue, excluding that this peptide carries an *O*-glycosylation instead. Similarly, in **B**) the phosphorylation clearly is associated with the hexose residues, excluding that is rather on the Ser or Thr residue. The glycoprotein and specific site harbouring the modification is shown on the top of the spectrum.

Neutrophil granules are known to harbor a wide selection of proteases to facilitate antimicrobial activity, such as neutrophil elastase, myeloblastin and cathepsin G. This means that a wider variety of peptides is possible for whole-neutrophil samples than can be generated with just GluC and trypsin²⁶. To account for this, we additionally searched the proteome with no constraint on the peptide sequence, which allowed the detection of 5444 proteins (**Figure 3**)(**Supplementary Table S2**).

Glycoproteomics

We subjected the neutrophil (glyco)peptides resulting from both the desalting-only and desalting+HILIC enrichments to triplicate RP-LC-MS² analysis with dedicated glycoproteomics parameters. This meant that precursor signals were either fragmented by higher-energy collisional dissociation (HCD), or by electron-transfer higher-energy collisional dissociation (EThcD) triggered on glycan fragments from preceding HCD fragmentation. In doing so, we obtained clear MS² evidence for the presence of high mannose glycans, sialylated glycans, as well as high degrees of antennary fucosylation across the glycoproteome (**Figure 4**). Importantly, we could also establish the presence of the paucimannose and phosphomannose-containing glycan species (**Figure 4 A,B**).

Glycoproteomics data were searched for the presence of glycopeptides using Byonic²⁷. We opted to perform the search without constrain on proteolytic cleavage, considering our observations from the shotgun proteomics experiment and the evidence that the structure of a glycan may be of large influence to proteolytic efficiency as well¹⁴. To retain a manageable search space we filtered our databases for those protein sequences that contained at least one *N*-glycosylation sequon (Asn-Xxx-Ser/Thr; Xxx ≠ Pro). Doing so, we searched one instance of each sample type (RP-only HCD, RP-only EThcD, RP+HILIC EThcD) against the full human glycoproteome (Swiss-Prot), and performed searches against the triplicates with a shortlist of the top 500 glycoproteins from the shotgun results. The majority of peptides proved to be semispecific for GluC/trypsin digestion, with common alternative cleavage sites occurring at Val, Phe, Leu/Ile, and Tyr (**Supplementary Figure S1**).

In terms of glycosylation we included 279 *N*-glycan compositions in the search (**Supplementary Table S3**), which encompassed paucimannose species and smaller (*e.g.*,

HexNAc₁dHex₁ and HexNAc₂Hex₂), monoantennary and hybrid species (e.g., HexNAc₃Hex₃NeuAc₁ and HexNAc₃Hex₆), complex multiantennary species with various degrees of fucosylation (e.g., HexNAc₄Hex₅NeuAc₂ and HexNAc₆Hex₇dHex₅NeuAc₁), and phosphomannosylated compositions typically originating from the lysosomal pathway of protein degradation (e.g., HexNAc₂Hex₅Phospho₁ and HexNAc₄Hex₈Phospho₂)^{23,28}. Following previously reported data curation parameters²⁹, we retained a total of 2726 unique *N*-glycopeptide identifications (**Supplementary Table S4**), informing on the glycosylation of 241 neutrophil proteins with at least one *N*-glycosylation sequon. We then assigned global glycosylation characteristics to the glycopeptides on the basis of monosaccharide compositions. These comprised **1**) unoccupied (no monosaccharides on a given site), **2**) paucimannose (HexNAc < 3 and Hex < 4), **3**) phosphomannose (Phospho > 0), **4**) high-mannose (HexNAc = 2 and Hex > 3), **5**) hybrid/asymmetric (HexNAc = 3), **6**) diantennary (HexNAc = 4) and **7**) extended (HexNAc > 4).

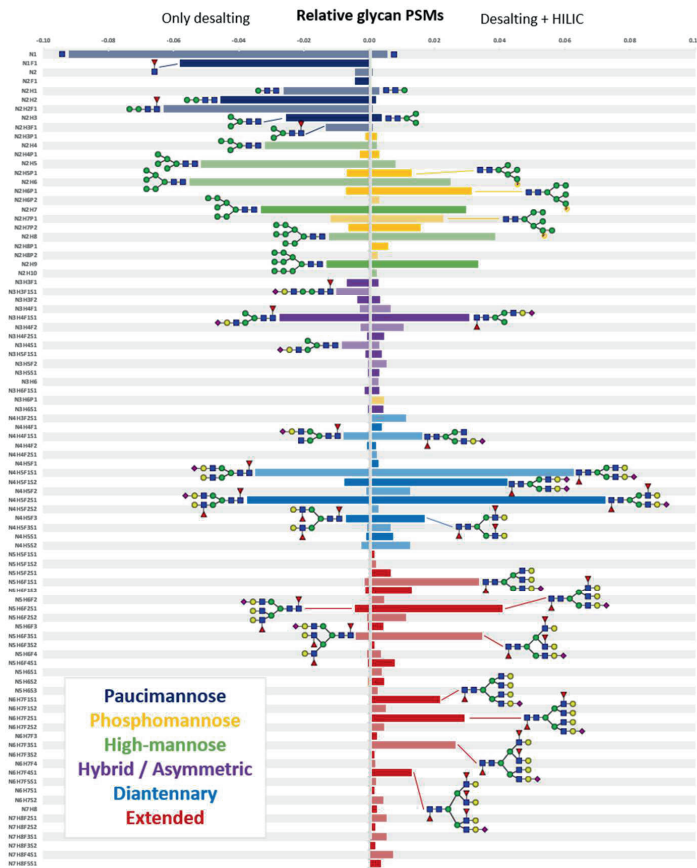


Figure 5: Bias in glycopeptide identifications. Quantitative overview of the peptide glycoforms detected across the 500 most abundant neutrophil glycoproteins. **Left)** The relative number of glycopeptides detected by MS (PSMs) directly after desalting. **Right)** The relative number of glycan PSMs detected after sequential desalting and HILIC enrichment. As seen from the distributions, HILIC enrichment after desalting leads to a bias against the smallest glycan species but allows for better detection of the lower-abundant higher-complexity glycans. As such, we considered desalting to be dominant in reporting the glycan distribution of a given glycosylation site but utilized the HILIC data when little or no RP detections were achieved. Colours indicate the classes of the glycans. Shorthand compositional annotation on the y-axis follows N = *N*-acetylhexosamine (*e.g.*, *N*-acetylglucosamine), H = hexose (*e.g.*, mannose or galactose), F = deoxyhexose (fucose), S = sialic acid (*N*-acetylneuraminic acid), and P = phosphorylation (*e.g.*, mannose-6-phosphate).

Importantly, the desalting-only and desalting+HILIC methods yielded information on different parts of the neutrophil glycoproteome. As witnessed from the distribution of peptide spectrum matches (PSMs; the measure of fragmentation spectra that have likely been correctly assigned), the RP-only method facilitated the detection of paucimannose species across the proteome, whereas the corresponding peptides were almost fully lost with subsequent HILIC enrichment (**Figure 5**). The additional HILIC step, on the other hand, altogether improved the detection glycopeptides so that the more analytically-challenging higher antennarities became quantifiable. Due to the bias observed in RP+HILIC we considered the RP method to be leading in the detection of peptide glycoforms. However, as the RP+HILIC method provided more information on the (individually) lower abundant high-complexity glycoforms, we opted to retain both methods for analysis and visualization.

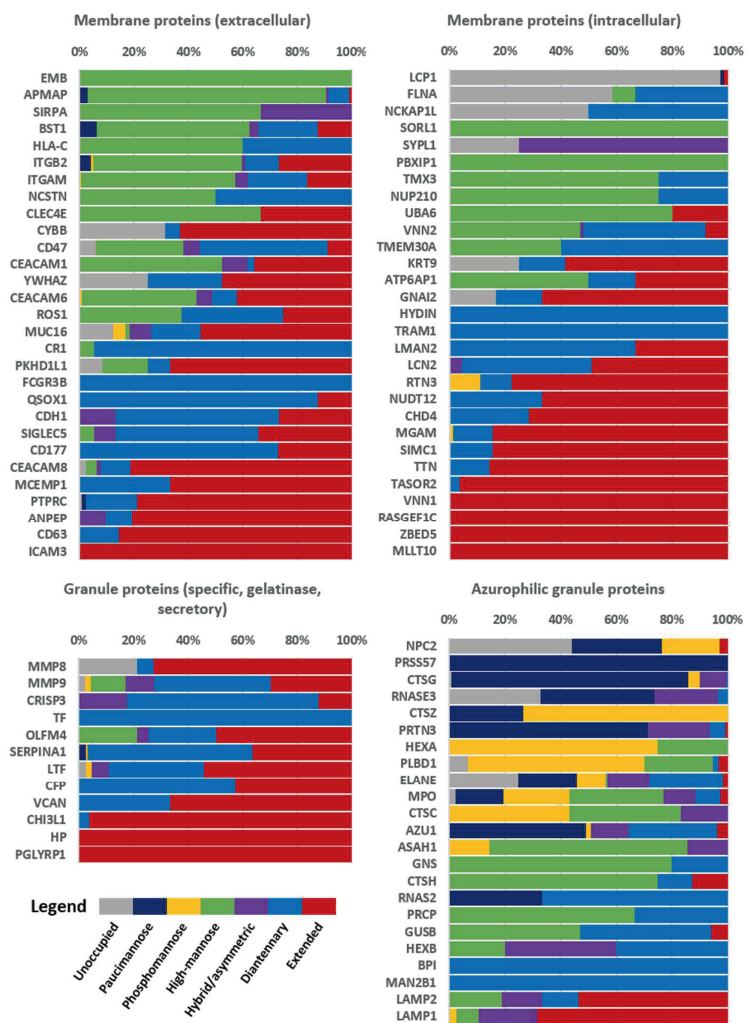


Figure 6: Individual glycoprotein characteristics as percentage of the total of glycopeptide detections. Proteins were only included when detected by at least 3 glycopeptide PSMs (thereby eliminating non-glycosylated proteins). Classification between proteins was made on the cellular location as described in the literature and neXtprot (www.nextprot.org). Proteins indicate as “membrane proteins” typically included a transmembrane domain, but this does not necessitate that they always exist membrane-bound. Most intracellular membrane proteins were described as localised to the endoplasmic reticulum or Golgi apparatus, whereas cytosolic, nuclear and mitochondrial proteins typically did not carry *N*-glycans. Whereas azurophilic (primary) granule proteins are listed separately, the other secreted proteins included proteins from the specific (secondary) granules, gelatinase (tertiary) granules and secretory vesicles. For each location, the proteins were sorted based on their glycosylation complexity (following the order of the legend). As evident from the data, the uncommon glycosylation characteristics of pauci- and phosphomannosylation appeared to be rather unique features of azurophilic granule proteins.

In general, neutrophil proteins expressed high-mannose, diantennary or extended-type glycosylation (**Figure 6**). By inferring the likely cellular locations for these proteins from the literature, we could predict their origins to be from the cell surface, from within the cell (including in the endoplasmic reticulum and Golgi apparatus), from specific and gelatinase granules, and from secretory vesicles⁵. Proteins to carry these features included lactotransferrin, neutrophil gelatinase-associated lipocalin, integrin alpha-M (CD11b), integrin beta-2 (CD18), carcinoembryonic antigen-related cell adhesion molecule 8 (CD66b), olfactomedin-4, and many others. Interestingly, for proteins predicted to stem from the azurophilic granules we could detect the abundant presence of very different glycosylation characteristics, namely of partial site-unoccupancy, paucimannose species, hybrid-/asymmetric-type glycans, as well as phosphomannosylation. As such, proteins to carry these unusual glycosylation features included, next to others, neutrophil elastase, cathepsin G, myeloblastin (proteinase 3), azurocidin and myeloperoxidase.

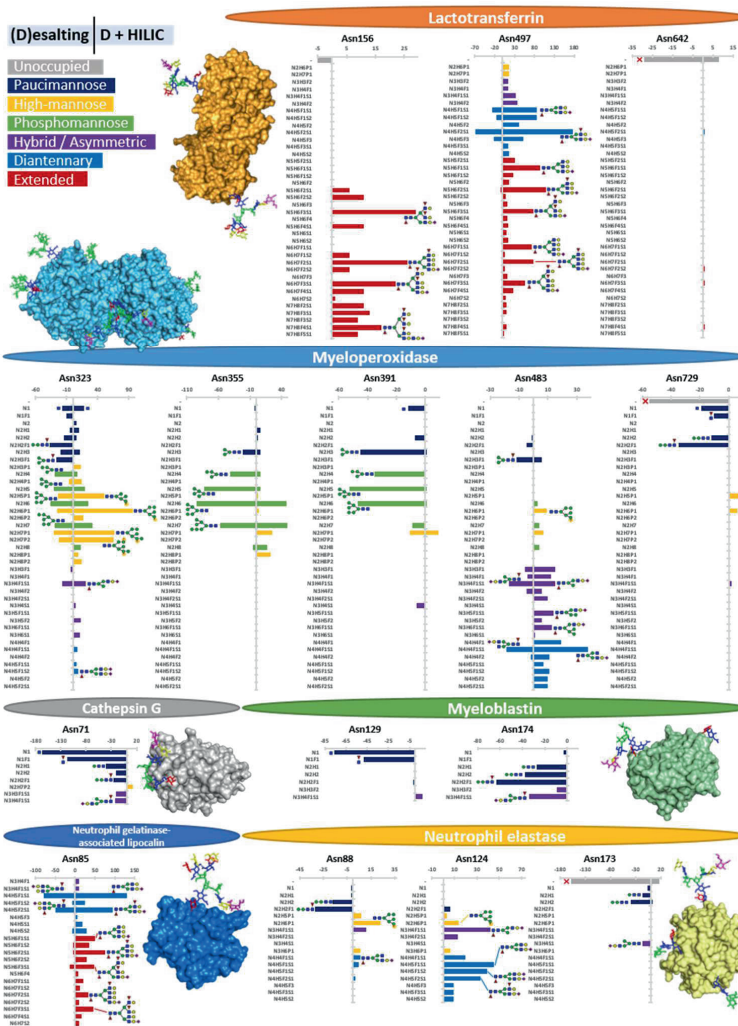


Figure 7: Site-specific glycosylation detected on six major neutrophil proteins. Glycan species were only included in the comparison when at least nine glycopeptide PSMs were detected for any glycosylation site of that protein in either the desalting-only condition (left) or desalting followed by HILIC (right). Glycoprotein structures were generated by energy-minimization of the glycan on an existing PDB structure by GLYCAM (www.glycam.org). As mass spectrometry does not principally distinguish glycan isomerism, the proposed glycan structures are estimates on basis of the literature. As can be seen, while proteins exhibited overall glycosylation characteristics such as high complexity (lactotransferrin, neutrophil gelatinase-associated lipocalin) or low complexity (myeloperoxidase, cathepsin G, myeloblastin, neutrophil elastase), they also showed considerable site-dependent heterogeneity within the same protein.

Site-specific glycan heterogeneity

By comparing relative PSM numbers, which for our myeloperoxidase benchmark yielded highly similar quantification as manual MS¹ area integration (**Supplementary Figure S2**), we assessed the glycan distribution of each *N*-glycosylation site for the major neutrophil proteins (**Figure 7**). Interestingly, each individual glycosylation site displayed its own unique pattern. For instance, lactotransferrin carried on its Asn156 exclusively extended glycans (tri- and tetraantennary species with a high degree of antennary fucosylation), while Asn497 expressed diantennary and smaller glycans and Asn642 remained mostly unoccupied. High-mannose and diantennary/extended species were typically detected on different sites. On integrin alpha-M (CD11b), for instance, we could primarily detect high-mannose species at sites Asn391, Asn734, Asn801, Asn880, Asn900 (also hybrid), Asn940, Asn978, Asn1021, Asn1044 and Asn1050, while we could mainly detect diantennary/extended species at sites Asn240, Asn692, Asn993, and Asn1075. Another example of this site-specific complexity could be found in integrin beta-2 (CD18), on which high-mannose species mainly covered sites Asn501 and Asn636, while diantennary/extended species occupied Asn50, Asn116, and Asn212.

Of interest, for the azurophilic granule proteins we detected altogether different glycosylation characteristics than for the other neutrophil locations, but we could still appreciate the large heterogeneity between individual sites. For example, on myeloperoxidase we detected phospho- and high-mannose glycosylation at Asn323, pauci- to high-mannose species at both Asn355 and Asn391, hybrid/asymmetric/diantennary species at Asn483, and paucimannose glycans together with partial occupancy at Asn729. Another example, neutrophil elastase, exhibited pauci- and phosphomannosylation at Asn88 and diantennary species at Asn124, while Asn173 remained mostly unoccupied. Even for proteins that generally displayed the same glycosylation characteristics across sites, *e.g.*, myeloblastin (proteinase 3) which was primarily paucimannosylated, we still could detect clear differences in which glycan composition was dominant, *e.g.*, HexNAc₁dHex₁ at Asn129 and HexNAc₂Hex₂dHex₁ at Asn174.

Discussion

On the one hand, neutrophils are the critical first responders in the immune system, on the other, the cells are also responsible for much off-target damage and the production of proteins that may end up being the target of autoantibodies⁹⁻¹¹. Given this, it is important to understand what the involved proteins actually look like, not only in terms of amino acid sequence, but also in terms of their modifications. However, even while the PTMs may have significant influence on enzymatic activity, interaction with inhibitors and scavengers, and the overall accessibility of immunogenic epitopes, neutrophil protein PTMs remain a vastly understudied subject. Several reports, including our own, have suggested that released neutrophil proteins may even carry *N*-

glycosylation variants that are rarely observed on extracellular proteins within humans, namely paucimannosylation and phosphomannosylation^{16-19,21}. As such, our goals in the current study were to **1)** verify the atypical glycosylation to originate from neutrophil proteins, **2)** to establish on which proteins and protein groups these can be detected, and **3)** to explore the general glycosylation characteristics of neutrophil proteins in a site-specific manner. By performing mass spectrometric (glyco)proteomics on the neutrophils from ten apparently healthy donors, we here provide the first in-depth investigation into the site-specific glycosylation characteristics of the neutrophil proteome.

Observed glycosylation characteristics

On the putatively membrane-bound proteins, both on the cell surface and within the cell, we found a combination of sites containing diantennary/extended glycans and those with high-mannose glycans. Individual glycan sites may certainly have unique functions, but it is interesting to see that a large part of the abundant surface receptors is occupied by fucosylated and sialylated glycans. Although our mass spectrometric analysis only yielded monosaccharide compositions and not the linkage between these, it seems likely that the monosaccharides combine into Lewis X (Gal β 1,4-[Fuc α 1,3-]GlcNAc) and sialyl-Lewis X (Neu5Ac2,3-Gal β 1,4-[Fuc α 1,3-]GlcNAc), respectively also named CD15 and CD15s. These carbohydrate epitopes are known to be abundantly present on neutrophils, and can, for instance, be used as a marker for fluorescence-assisted cell sorting (FACS) (our gating strategy selected on CD15)^{30,31}. Biologically speaking, sialyl-Lewis X can interact with the E-selectins that are commonly expressed at sites of injury and inflammation. Interactions between E-selectins and neutrophil sialyl-Lewis X motifs may then facilitate the typical tethering behaviour to catch and slow down the cell from the blood stream and start the extravasation process^{30,32}.

The proteins likely secreted from specific (secondary) granules, gelatinase (tertiary) granules and other secretory vesicles showed diantennary and extended species as well, but only limited occupancy of high-mannose type glycans. One could argue that high-mannose glycans are therefore a feature of membrane proteins, or proteins with a membrane-bound origin. Interestingly, amongst the secretory proteins we also observed several that have been studied in a variety of biofluids³³⁻³⁵. Alpha-1-antitrypsin (A1AT), for example, is generally a hepatocyte-produced acute phase protein that is abundantly present in blood³³. It is then remarkable to see that our neutrophil-sourced A1AT shows highly-similar glycosylation patterns compared to the one from blood, namely highly sialylated (and to a lesser degree fucosylated) di- and triantennary glycans on sites Asn70, Asn107 and Asn271^{34,36}. Similarly, lactotransferrin is a major component of human milk, and the high degrees of antennary

fucosylation we see on both Asn156 and Asn497 and lack of occupancy at Asn642 matches between sources as well³⁵. While it cannot be excluded that, for instance, alpha-1-antitrypsin was co-captured during the flow cytometry of our neutrophils, an alternative explanation is that the glycosylation status of secretory proteins is highly dependent on the structure of the proteins themselves.

Importantly, the proteins that likely originated from the azurophilic (primary) granules demonstrated altogether different glycosylation characteristics, with sites containing partial unoccupancy, pauci- and phosphomannose glycans, as well as a large degree of hybrid/asymmetric glycans. Not only could this be confirmed for the azurophilic granule proteins that were prior studied in isolated fashion, *i.e.*, myeloperoxidase²¹, neutrophil elastase¹⁶, cathepsin G¹⁷, and myeloblastin to some degree¹⁸, but this can now be expanded to many others including azurocidin, cathepsins C and Z, beta-hexosaminidase subunits alpha and beta, phospholipase B-like 1, and so forth. The glycosidases found amongst the azurophilic granule proteins may also provide a partial explanation for the many paucimannose glycans we observed. Lysosomal alpha-mannosidase (MAN2B1) can cause the removal of the alpha-linked mannoses that occupy most high-mannose and hybrid glycans, whereas the beta-hexosaminidases (HEXA, HEXB) can attack the antennary and core *N*-acetylglucosamines. If these glycosidases are indeed responsible for the general glycan truncation amongst azurophilic granule proteins, the question instead becomes why certain glycosylation sites remain unaffected. Myeloperoxidase is an intriguing example of this, with high-mannose glycans on sites Asn355 and Asn391 but severely truncated glycans at Asn729, while myeloblastin shows distinct paucimannose sizes at Asn129 (monosaccharide) and Asn174 (pentasaccharide). Protection of a glycosylation site could be achieved by steric hindrance, strong interactions between glycan and peptide backbone (potentially of an oligomerization partner), or perhaps the glycans are protected by dedicated chaperones. In all cases, however, the presence of larger glycans in a glycosidase-rich environment reflects an additional level of proteoform regulation.

The other atypical glycosylation characteristic, *i.e.*, phosphomannosylation, may have something to do with the trafficking function it has in other cell types^{23,37,38}. Most cells can target glycoproteins for degradation by phosphorylation of high-mannose glycans, which are then subsequently recognized by mannose-6-phosphate receptors and translocated towards lysosomes for degradation^{37,38}. Although neutrophils do not have typical lysosomes, instead having repurposed them into azurophilic granules (with some distinct differences)⁶, it may be that the lysosomal trafficking can still be used to tune granule contents. The dominant theory for populating different granules lies in transcription-based timing, with translated proteins progressively (in time) ending up in azurophilic (primary), specific (secondary) and gelatinase (tertiary) granules^{5,39}. However, a shuttling mechanism akin to lysosomal trafficking might still

allow to fill particular granules even when the dedicated time-frame has passed. While speculative, this would be an interesting lead to follow-up. Alternatively, both pauci- and phosphomannosylation may assist in the clearance of the hazardous azurophilic granule proteins from the extracellular space. For instance, part of the mannose-6-phosphate receptors can be found on cell surfaces and may mediate internalization of phosphomannosylated proteins for lysosomal degradation^{38,40}. Similarly, other carbohydrate receptors may be instrumental in clearing paucimannose species from the extracellular space. If not cleared, however, both pauci- and phosphomannose glycans could present interesting targets for intervention in neutrophil-based malignancies.

Glycoproteomics considerations

Several methodological aspects of our study need to be kept in mind when interpreting the results. Given that neutrophils abundantly produce proteases like neutrophil elastase, cathepsin G and myeloblastin, it was no surprise that our neutrophil proteomics study displayed many semispecific and nonspecific peptides. However, while for the relative quantification of glycan distribution per glycosylation site we would typically proceed with manual integration of MS¹ areas for individual glycopeptides (*e.g.*, by Skyline)^{21,35}, the extreme number of glycan and peptide combinations prohibited this approach on any practical level. On the other hand, a relevant alternative form of quantification may come from the counting of PSMs (annotated MS²)^{23,41}. Validating this approach on the data we previously generated for the in-depth profiling of isolated neutrophil myeloperoxidase²¹, we were intrigued to see an excellent approximation of relative glycan distribution per site (**Supplementary Figure 2**). It can be reasoned that, compared to low-abundant analytes, high-abundant analytes will have longer elution times, more detectable charge states, more detectable structural isomers, more detectable peptide miscleavages, a higher chance of MS² triggering, as well as a higher chance of identification when triggered, each of these contributing to a higher number of PSMs. While not a full substitute for manual integration, PSM counting (MS²-based quantification) may provide a reasonable alternative for complex glycoproteomics samples.

Sample preparation with either desalting-only or desalting+HILIC presented notably different views of the neutrophil glycoproteome. HILIC is a commonly used for the enrichment of glycans and glycopeptides and its retention mechanisms is primarily based on capturing the glycan portion of a glycoconjugate while RP is typically dependent to the peptide portion⁴². As such, our number of glycopeptide PSMs was far greater in the HILIC method than in RP-only. This also led to the clear detection of analytically complex glycans which might otherwise be missed, *e.g.*, due to splitting of signal intensity across related glycans and glycan stereoisomers.

On the other hand, HILIC enrichment also led to the major loss of glycopeptides with paucimannose glycans and smaller, which is understandable because HILIC utilizes the carbohydrate portion for its retention mechanism. While we here describe paucimannose glycosylation to be “atypical” and “uncommonly observed”, this might also be due to a general underestimation across the literature due to the ubiquity of HILIC enrichment for glycan analysis^{42,43}. In the work presented here, we interpreted RP as dominant for informing on the the glycan distribution for a given glycosylation site, but relied on HILIC when little or no RP detections were achieved.

Regarding our assignment of glycan compositions, convincing fragmentation patterns were obtained to prove the presence of both multifucoylated and sialylated glycan species. It has to be noted, however, that the mass increment of a sialic acid (+291) is very similar to that of two fucoses (+292), and these can easily be co-isolated for MS². While glycan species with only sialic acids or only fucoses are simple to assign by fragmentation, ratios of sialylation/fucoylation remain challenging to establish from fragment abundances alone. Similarly, we expect our level of sialylation to be somewhat underestimated by the capture of ammonium, or other cations, that dilute their signals and may prevent their observation. Furthermore, the term “extended” may either be interpreted as tri- or tetraantenary glycosylation or as existing antennae being extended by *N*-acetylglucosamine (LacNAc) repeats. As gaining an antenna is the same mass increment as LacNAc extension, and fragmentation of either typically results in the same oxonium ions (*e.g.*, m/z 366 [M+H]⁺), we did not make a distinction between them. For the most accurate mass spectrometric quantification of protein glycosylation, we highly recommend to isolate single proteins and match the findings from bottom-up MS with intact/native MS as detailed previously⁴⁴.

Conclusions

To conclude, we present here the first comprehensive investigation of the neutrophil glycoproteome and report on the site-specific glycosylation of more than 200 individual proteins. Next to characterizing the glycans on membrane-bound and secreted proteins, we report the remarkable carbohydrate characteristics for the proteins that typically reside in the azurophilic granules. We believe that this report will aid our understanding of neutrophil (glyco)biology, and that it establishes important leads for future research into the mechanisms and potential treatment of neutrophil-based malignancies.

Methods

Chemicals and materials

Unless otherwise specified, all chemicals and reagents were obtained from Sigma-Aldrich (Steinheim, Germany). GluC was obtained from Roche (Indianapolis, IN). The Oasis PRiME HLB plate was purchased from Waters (Etten-Leur, the Netherlands). Formic acid (FA) was purchased from Merck (Darmstadt, Germany). Acetonitrile (ACN) was purchased from Biosolve (Valkenswaard, The Netherlands). Milli-Q was produced by an in-house system (Millipore, Billerica, MA), phosphoSTOP (Roche, Woerden, the Netherlands) and complete mini EDTA free (Roche, Woerden, the Netherlands).

Isolation of human neutrophils

Neutrophil collection from buffy coats was performed as previously described^{48,49}.

Cell lysis and protein digestion

Neutrophil cell pellets (approximately $3 \cdot 10^7$ cells) were resuspended in lysis buffer containing 100 mM Tris-HCl (pH 8.5), 7 M urea, 5 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP), 30 mM chloroacetamide (CAA), Triton X-100 (1%), 2 mM magnesium sulfate, phosphoSTOP and complete mini EDTA free. Then, cells were disrupted by sonication for 45 min (alternating 20 s on and 40 s off) using a Bioruptor Plus (Diagenode, Seraing, Belgium). Cell debris was removed by centrifugation at 14000 rpm for 1 h at 4 °C and the supernatant retained. Impurities were removed by methanol/chloroform protein precipitation as follows: 1 mL of supernatant was mixed with 4 mL of methanol, 1 mL chloroform and 3 mL ultrapure water with thorough vortexing after each addition. The mixture was then centrifuged for 10 min at 5000 rpm at room temperature (RT). The upper layer was discarded, and 3 mL of methanol was added. After sonication and centrifugation (5000 rpm, 10 min at RT), the solvent was removed, and the precipitate was allowed to dry by air. The pellet was resuspended in digestion buffer containing 100 mM Tris-HCl (pH 8.5), 1% w/v sodium deoxycholate (SDC), 5 mM TCEP and 30 mM CAA. GluC was then added to digest proteins for 3 h at an enzyme-to-protein ratio of 1:75 (w/w) at 37 °C, and the resulting peptide mixtures were further digested overnight at 37 °C by trypsin (1:20; w/w). The next day, SDC was removed via acid precipitation (0.5% trifluoroacetic acid) (TFA) and the final peptide concentration was estimated by measuring the absorbance at 280 nm on a Nanodrop Spectrophotometer (Nanodrop 2000, Thermo Scientific). The peptides were desalted by using an Oasis PRiME HLB plate then dried and stored at -80 °C.

HILIC-based glycopeptide enrichment

The HILIC-based glycopeptide enrichment was performed using in-house packed stage-tips. In short, 200 μL pipet tips were packed with ten mg of 3 μm ZIC-cHILIC beads (Merk, Darmstadt, Germany), which were then washed with 100 μL of 1% FA and equilibrated with 100 μL of loading buffer (80% ACN/0.5% TFA). The peptide digests (100 μg) were reconstituted with 100 μL of loading buffer and loaded onto the stage-tips. The stage-tips were washed with 100 μL of loading buffer, and the glycopeptides were first eluted with 65% ACN/0.5% TFA, followed by 55% ACN/0.5% TFA. The elution were dried down and stored at $-80\text{ }^{\circ}\text{C}$ until subjected to LC-MS².

Shotgun proteomics

Shotgun LC-MS² was performed by means of an Agilent 1290 Infinity HPLC system (Agilent Technologies, Waldbronn, Germany) coupled to a Q Exactive HF mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). Per run, approximately 800 ng of peptides were first trapped by using a 100 μm inner diameter 2 cm trap column (in-house packed with ReproSil-Pur C18-AQ, 3 μm) (Dr. Maisch GmbH, Ammerbuch-Entringen, Germany) coupled to a 50 μm inner diameter 50 cm analytical column (in-house packed with Poroshell 120 EC-C18, 2.7 μm) (Agilent Technologies, Amstelveen, The Netherlands). The mobile-phase solvent A consisted of 0.1% FA in water, and the mobile-phase solvent B consisted of 0.1% FA in ACN. Trapping was performed at a flow rate of 5 $\mu\text{L}/\text{min}$ for 5 min with 0% B and peptides were eluted using a passively split flow of 300 nL/min for 170 min with 10% to 36% B over 155 min, 36% to 100% B over 3 min, 100% B for 1 min, 100% to 0% B over 1 min, and finally held at 0% B for 10 min. Peptides were ionized using a spray voltage of 1.9 kV and a heated capillary. The mass spectrometer was set to acquire full-scan MS spectra (m/z 375-1600) for a maximum injection time of 20 ms at a mass resolution of 60,000 and an automated gain control (AGC) target value of 3×10^6 . Up to 15 of the most intense precursor ions were selected for tandem mass spectrometry (MS²). HCD MS² (m/z 200-2000) acquisition was performed in the HCD cell, with the readout in the Orbitrap mass analyzer at a resolution of 30,000 (isolation window of 1.4 Th) and an AGC target value of 1×10^5 or a maximum injection time of 50 ms with a normalized collision energy of 27%.

Glycoproteomics

All peptides from the (glyco)peptide enrichments were separated and analysed using the same HPLC system as used for the global proteome analysis, albeit now coupled on-line to an Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) using a 90 min gradient, as follows: 0-5 min, 100% solvent A; 13-44% solvent B for 65 min; 44-

100% solvent B for 5 min; 100% solvent B for 5 min; 100% solvent A for 15 min. Per run, approximately 300 ng (nonenriched) or 100 μ g (HILIC enriched) peptides were ionized using a 2.0 kV spray voltage. For the MS scan, the mass range was set from m/z 350 to 2000 with a maximum injection time of 50 ms at a mass resolution of 60,000 and an AGC target value of 5×10^4 in the Orbitrap mass analyzer. The dynamic exclusion was set to 30 s for an exclusion window of 10 ppm with a cycle time of 3 s. Charge-states screening was enabled, and precursors with 2+ to 8+ charge states and intensities $> 1e5$ were selected for MS². HCD MS² (m/z 120–4000) acquisition was performed in the HCD cell, with the readout in the Orbitrap mass analyzer at a resolution of 30,000 (isolation window of 2 Th) and an AGC target value of 5×10^4 or a maximum injection time of 50 ms with a normalized collision energy of 27%. If at least 2 out of 4 oxonium ions of glycopeptides (138.0545+, 204.0687+, 366.1396+ or 243.026+) were observed²⁴, EThcD MS² on the same precursor was triggered (isolation window of 3 Th) and fragment ions (m/z 120–4000) were analyzed in the Orbitrap mass analyzer at a resolution of 30,000, AGC target value of 2×10^5 or a maximum injection time of 250 ms with activation of ETD and supplemental activation with a normalized collision energy (NCE) of 27%.

Proteomics data analysis

Proteome Discoverer (version 2.3, Thermo Scientific) was used to analyze the LC-MS² raw files from the shotgun experiment. MS² scans were searched against the Swiss-Prot database (release date: Sep 2019, 20 421 entries, taxonomy: Homo sapiens) by using Sequest with the following modifications: fixed Cys carbamidomethylation, variable Met oxidation and Ser, Thr, and Tyr phosphorylation (serine, threonine and tyrosine). Enzyme specificity was set to trypsin and GluC with a maximum of 3 missed cleavages for peptides in a fully specific digestion, and set to “none” for peptides in nonspecific digestion. The searches were performed using a precursor mass tolerance of 10 ppm and a fragment mass tolerance of 0.05 Da followed by data filtering using Percolator, resulting in a 1% false discovery rate (FDR). For label-free quantification, the node called Minora feature detector was used with high PSM confidence, a minimum of 5 non-zero points in a chromatographic trace, a minimum number of 2 isotopes and maximum retention time difference of 0.2 min for isotope peaks. The consensus workflow in Proteome Discoverer was used to open the search results and enable retention time (RT) alignment with a maximum RT shift of 5 min and mass tolerance of 10 ppm in order to match the precursor between runs. Proteins were only reported when detected in at least two out of three replicates.

Glycoproteomics data analysis

The raw files acquired from the glycoproteomics experiments were searched with Byonic (version 3.7.13, Protein Metrics Inc.)²⁷ against all potential glycoproteins (sequences containing Asn-Xxx-Ser/Thr where Xxx ≠ Pro) from the Swiss-Prot database (one replicate) as well as against a focused database comprising the top 500 most abundant glycoproteins from the neutrophil proteome analysis (all replicates). The precursor ion mass tolerance was set to 10 ppm and both HCD & EThcD fragment mass tolerance were set to 20 ppm. Peptides were searched with nonspecific proteolytic digestion, Cys carbamidomethylation as fixed modification, and up to two variable modifications of Met oxidation and Ser/Thr/His/Tyr phosphorylation. For glycosylation we allowed one glycan from a database of 279, which included highly fucosylated species and phosphomannose/degraded glycans known from the lysosomal pathway of degradation (**Supplementary Table S1**). Based on prior established curation criteria²⁹, we rejected glycopeptide identifications with a log probability lower than 1, and maintained a score threshold of 150. PSMs of individual glycan compositions were counted using an in-house script and relatively quantified by total area normalization of the sum of counts, either per glycosylation site (**Figure 7**), per glycoprotein (**Figure 6**), or across the whole glycoproteome (**Figure 5**).

Data visualization

For visualization of glycan species we followed the recommendations of the Consortium for Functional Glycomics⁴⁵, while glycan cartoons were constructed using Glycoworkbench 2.1 (build 146)⁴⁶. Glycoprotein structures were generated by GLYCAM⁴⁷ using the following PDB structures: lactotransferrin (1N76), myeloperoxidase (1CXP), cathepsin G (1AU8), myeloblastin (one chain of 1FUJ), neutrophil gelatinase-associated lipocalin (1QQS), and neutrophil elastase (one chain of 3Q76). Glycan compositions were selected to be an abundant and representative species based on the glycoproteomics data, while linkages thereof were only estimated and may differ in actuality.

Acknowledgments

This work was supported by the Netherlands Organization for Scientific Research (NWO) TOP-Punt Grant 718.015.003 (to A.J.R.H.) and the European Union's Horizon 2020 Research and Innovation Program under Grant 668036 (RELENT). We further acknowledge funding for our large-scale proteomics facility, the Netherlands Proteomics Center, through the X-omics Road Map program (project 184.034.019) and the EU Horizon 2020 program INFRAIA project Epic-XS (Project 823839). K.R.R. further acknowledges support from NWO Veni project VI.Veni.192.058. Also, we acknowledge our cooperators Floris van Alphen and Sander B.

Meijer from *Department of Molecular and Cellular Hemostasis, Sanquin Research, Amsterdam, 1066 CX, The Netherlands* for providing neutrophil samples.

Competing interests

The authors report no conflict of interest.

References

- 1 Nathan, C. Neutrophils and immunity: challenges and opportunities. *Nat Rev Immunol* **6**, 173-182, doi:10.1038/nri1785 (2006).
- 2 Lekstrom-Himes, J. A. & Gallin, J. I. Immunodeficiency diseases caused by defects in phagocytes. *N Engl J Med* **343**, 1703-1714, doi:10.1056/NEJM200012073432307 (2000).
- 3 Pillay, J. *et al.* In vivo labeling with ²H₂O reveals a human neutrophil lifespan of 5.4 days. *Blood* **116**, 625-627, doi:10.1182/blood-2010-01-259028 (2010).
- 4 Hickey, M. J. & Kubes, P. Intravascular immunity: the host-pathogen encounter in blood vessels. *Nat Rev Immunol* **9**, 364-375, doi:10.1038/nri2532 (2009).
- 5 Rorvig, S., Ostergaard, O., Heegaard, N. H. & Borregaard, N. Proteome profiling of human neutrophil granule subsets, secretory vesicles, and cell membrane: correlation with transcriptome profiling of neutrophil precursors. *J Leukoc Biol* **94**, 711-721, doi:10.1189/jlb.1212619 (2013).
- 6 Cieutat, A. M. *et al.* Azurophilic granules of human neutrophilic leukocytes are deficient in lysosome-associated membrane proteins but retain the mannose 6-phosphate recognition marker. *Blood* **91**, 1044-1058 (1998).
- 7 Henson, P. M. & Johnston, R. B., Jr. Tissue injury in inflammation. Oxidants, proteinases, and cationic proteins. *J Clin Invest* **79**, 669-674, doi:10.1172/JCI112869 (1987).
- 8 Liu, J. *et al.* Advanced Role of Neutrophils in Common Respiratory Diseases. *J Immunol Res* **2017**, 6710278, doi:10.1155/2017/6710278 (2017).
- 9 Bossuyt, X. *et al.* Position paper: Revised 2017 international consensus on testing of ANCA in granulomatosis with polyangiitis and microscopic polyangiitis. *Nat Rev Rheumatol* **13**, 683-692, doi:10.1038/nrrheum.2017.140 (2017).
- 10 van der Geest, K. S. M. *et al.* Towards precision medicine in ANCA-associated vasculitis. *Rheumatology (Oxford)* **57**, 1332-1339, doi:10.1093/rheumatology/kex367 (2018).
- 11 Cornec, D., Cornec-Le Gall, E., Fervenza, F. C. & Specks, U. ANCA-associated vasculitis - clinical utility of using ANCA specificity to classify patients. *Nat Rev Rheumatol* **12**, 570-579, doi:10.1038/nrrheum.2016.123 (2016).
- 12 in *Essentials of Glycobiology* (eds rd *et al.*) (2015).
- 13 Yu, J. T. *et al.* Deglycosylation of myeloperoxidase uncovers its novel antigenicity. *Kidney Int* **91**, 1410-1419, doi:10.1016/j.kint.2016.12.012 (2017).

- 14 Falck, D. *et al.* Glycoforms of Immunoglobulin G Based Biopharmaceuticals Are Differentially Cleaved by Trypsin Due to the Glycoform Influence on Higher-Order Structure. *J Proteome Res* **14**, 4019-4028, doi:10.1021/acs.jproteome.5b00573 (2015).
- 15 Specks, U. *et al.* Functional significance of Asn-linked glycosylation of proteinase 3 for enzymatic activity, processing, targeting, and recognition by anti-neutrophil cytoplasmic antibodies. *J Biochem* **141**, 101-112, doi:10.1093/jb/mvm008 (2007).
- 16 Loke, I., Ostergaard, O., Heegaard, N. H. H., Packer, N. H. & Thaysen-Andersen, M. Paucimannose-Rich N-glycosylation of Spatiotemporally Regulated Human Neutrophil Elastase Modulates Its Immune Functions. *Mol Cell Proteomics* **16**, 1507-1527, doi:10.1074/mcp.M116.066746 (2017).
- 17 Loke, I., Packer, N. H. & Thaysen-Andersen, M. Complementary LC-MS/MS-Based N-Glycan, N-Glycopeptide, and Intact N-Glycoprotein Profiling Reveals Unconventional Asn71-Glycosylation of Human Neutrophil Cathepsin G. *Biomolecules* **5**, 1832-1854, doi:10.3390/biom5031832 (2015).
- 18 Zoega, M., Ravnsborg, T., Hojrup, P., Houen, G. & Schou, C. Proteinase 3 carries small unusual carbohydrates and associates with alpha-defensins. *J Proteomics* **75**, 1472-1485, doi:10.1016/j.jprot.2011.11.019 (2012).
- 19 Thaysen-Andersen, M. *et al.* Human neutrophils secrete bioactive paucimannosidic proteins from azurophilic granules into pathogen-infected sputum. *J Biol Chem* **290**, 8789-8802, doi:10.1074/jbc.M114.631622 (2015).
- 20 Tjondro, H. C., Loke, I., Chatterjee, S. & Thaysen-Andersen, M. Human protein paucimannosylation: cues from the eukaryotic kingdoms. *Biol Rev Camb Philos Soc* **94**, 2068-2100, doi:10.1111/brv.12548 (2019).
- 21 Reiding, K. R. *et al.* Neutrophil myeloperoxidase harbors distinct site-specific peculiarities in its glycosylation. *J Biol Chem* **294**, 20233-20245, doi:10.1074/jbc.RA119.011098 (2019).
- 22 Reiding, K. R., Bondt, A., Franc, V. & Heck, A. J. R. The benefits of hybrid fragmentation methods for glycoproteomics. *Trac-Trend Anal Chem* **108**, 260-268, doi:10.1016/j.trac.2018.09.007 (2018).
- 23 Caval, T. *et al.* Targeted Analysis of Lysosomal Directed Proteins and Their Sites of Mannose-6-phosphate Modification. *Mol Cell Proteomics* **18**, 16-27, doi:10.1074/mcp.RA118.000967 (2019).

- 24 Caval, T., Zhu, J. & Heck, A. J. R. Simply Extending the Mass Range in Electron Transfer Higher Energy Collisional Dissociation Increases Confidence in N-Glycopeptide Identification. *Anal Chem* **91**, 10401-10406, doi:10.1021/acs.analchem.9b02125 (2019).
- 25 Grabowski, P. *et al.* Proteome Analysis of Human Neutrophil Granulocytes From Patients With Monogenic Disease Using Data-independent Acquisition. *Mol Cell Proteomics* **18**, 760-772, doi:10.1074/mcp.RA118.001141 (2019).
- 26 Rieckmann, J. C. *et al.* Social network architecture of human immune cells unveiled by quantitative proteomics. *Nat Immunol* **18**, 583-593, doi:10.1038/ni.3693 (2017).
- 27 Bern, M., Kil, Y. J. & Becker, C. Byonic: advanced peptide and protein identification software. *Curr Protoc Bioinformatics* **Chapter 13**, Unit13 20, doi:10.1002/0471250953.bi1320s40 (2012).
- 28 Nairn, A. V. *et al.* Regulation of glycan structures in animal tissues: transcript profiling of glycan-related genes. *J Biol Chem* **283**, 17298-17313, doi:10.1074/jbc.M801964200 (2008).
- 29 Riley, N. M., Hebert, A. S., Westphall, M. S. & Coon, J. J. Capturing site-specific heterogeneity with large-scale N-glycoproteome analysis. *Nat Commun* **10**, 1311, doi:10.1038/s41467-019-09222-w (2019).
- 30 Varki, A. Selectin ligands. *Proc Natl Acad Sci U S A* **91**, 7390-7397, doi:10.1073/pnas.91.16.7390 (1994).
- 31 Graham, S. A. *et al.* Identification of neutrophil granule glycoproteins as Lewis(x)-containing ligands cleared by the scavenger receptor C-type lectin. *J Biol Chem* **286**, 24336-24349, doi:10.1074/jbc.M111.244772 (2011).
- 32 Aplin, A. E., Howe, A., Alahari, S. K. & Juliano, R. L. Signal transduction and signal modulation by cell adhesion receptors: the role of integrins, cadherins, immunoglobulin-cell adhesion molecules, and selectins. *Pharmacol Rev* **50**, 197-263 (1998).
- 33 Clerc, F. *et al.* Human plasma protein N-glycosylation. *Glycoconj J* **33**, 309-343, doi:10.1007/s10719-015-9626-2 (2016).
- 34 Totten, S. M., Feasley, C. L., Bermudez, A. & Pitteri, S. J. Parallel Comparison of N-Linked Glycopeptide Enrichment Techniques Reveals Extensive Glycoproteomic Analysis of Plasma Enabled by SAX-ERLIC. *J Proteome Res* **16**, 1249-1260, doi:10.1021/acs.jproteome.6b00849 (2017).

- 35 Zhu, J. *et al.* Quantitative Longitudinal Inventory of the N-Glycoproteome of Human Milk from a Single Donor Reveals the Highly Variable Repertoire and Dynamic Site-Specific Changes. *J Proteome Res*, doi:10.1021/acs.jproteome.9b00753 (2020).
- 36 Kolarich, D., Weber, A., Turecek, P. L., Schwarz, H. P. & Altmann, F. Comprehensive glyco-proteomic analysis of human alpha1-antitrypsin and its charge isoforms. *Proteomics* **6**, 3369-3380, doi:10.1002/pmic.200500751 (2006).
- 37 Dahms, N. M., Lobel, P. & Kornfeld, S. Mannose 6-phosphate receptors and lysosomal enzyme targeting. *J Biol Chem* **264**, 12115-12118 (1989).
- 38 Ghosh, P., Dahms, N. M. & Kornfeld, S. Mannose 6-phosphate receptors: new twists in the tale. *Nat Rev Mol Cell Biol* **4**, 202-212, doi:10.1038/nrm1050 (2003).
- 39 Le Cabec, V., Cowland, J. B., Calafat, J. & Borregaard, N. Targeting of proteins to granule subsets is determined by timing and not by sorting: The specific granule protein NGAL is localized to azurophil granules when expressed in HL-60 cells. *Proc Natl Acad Sci U S A* **93**, 6454-6457, doi:10.1073/pnas.93.13.6454 (1996).
- 40 Pohlmann, R. *et al.* Mannose 6-phosphate specific receptors: structure and function. *Biochem Soc Trans* **17**, 15-16, doi:10.1042/bst0170015 (1989).
- 41 Stadlmann, J. *et al.* Comparative glycoproteomics of stem cells identifies new players in ricin toxicity. *Nature* **549**, 538-542, doi:10.1038/nature24015 (2017).
- 42 Ruhaak, L. R., Xu, G., Li, Q., Goonatilleke, E. & Lebrilla, C. B. Mass Spectrometry Approaches to Glycomic and Glycoproteomic Analyses. *Chem Rev* **118**, 7886-7930, doi:10.1021/acs.chemrev.7b00732 (2018).
- 43 Gaunitz, S., Nagy, G., Pohl, N. L. & Novotny, M. V. Recent Advances in the Analysis of Complex Glycoproteins. *Anal Chem* **89**, 389-413, doi:10.1021/acs.analchem.6b04343 (2017).
- 44 Yang, Y. *et al.* Hybrid mass spectrometry approaches in glycoprotein analysis and their usage in scoring biosimilarity. *Nat Commun* **7**, 13397, doi:10.1038/ncomms13397 (2016).
- 45 Varki, A. *et al.* Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology* **25**, 1323-1324, doi:10.1093/glycob/cwv091 (2015).
- 46 Ceroni, A. *et al.* GlycoWorkbench: a tool for the computer-assisted annotation of mass spectra of glycans. *J Proteome Res* **7**, 1650-1659, doi:10.1021/pr7008252 (2008).
- 47 Kirschner, K. N. *et al.* GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J Comput Chem* **29**, 622-655, doi:10.1002/jcc.20820 (2008).

48 Reumaux, D.; de Boer, M.; Meijer, A. B.; Duthilleul, P.; Roos, D. Expression of Myeloperoxidase (MPO) by Neutrophils Is Necessary for Their Activation by Anti-Neutrophil Cytoplasm Autoantibodies (ANCA) against MPO. *J. Leukoc. Biol.* 2003, 73 (6), 841–849.

49 Hoogendijk, A. J.; Pourfarzad, F.; Aarts, C. E. M.; Tool, A. T. J.; Hiemstra, I. H.; Grassi, L.; Frontini, M.; Meijer, A. B.; van den Biggelaar, M.; Kuijpers, T. W. Dynamic Transcriptome-Proteome Correlation Networks Reveal Human Myeloid Differentiation and Neutrophil-Specific Programming. *Cell Rep.* 2019, 29 (8), 2505-2519.e4.

Supporting information for Chapter 6

Supplementary Figures

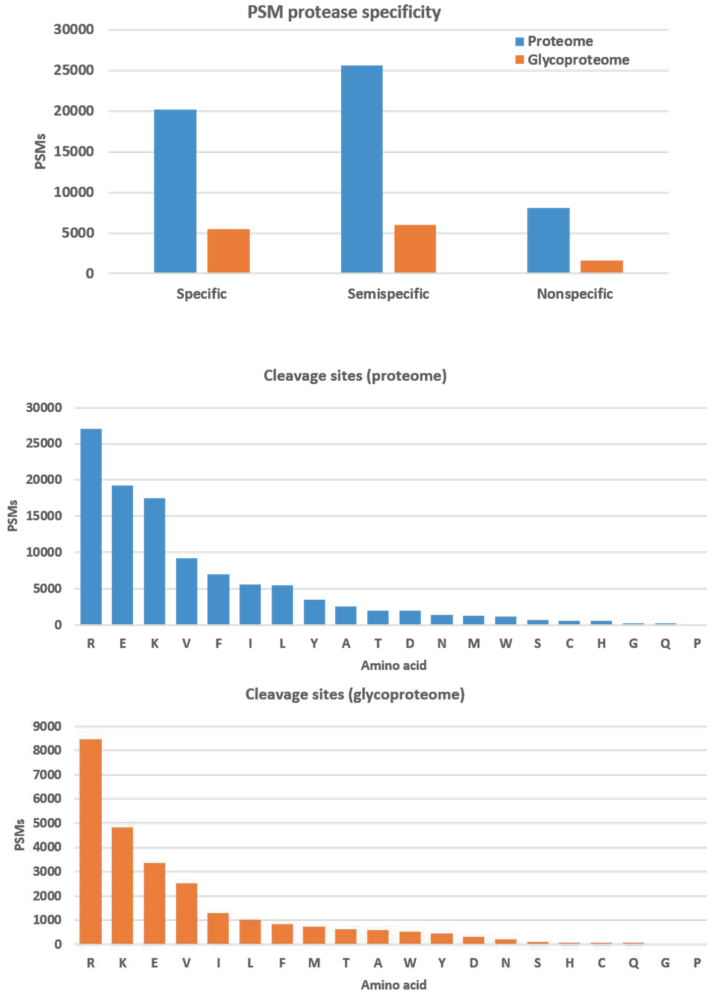


Figure S1: Protease specificity.

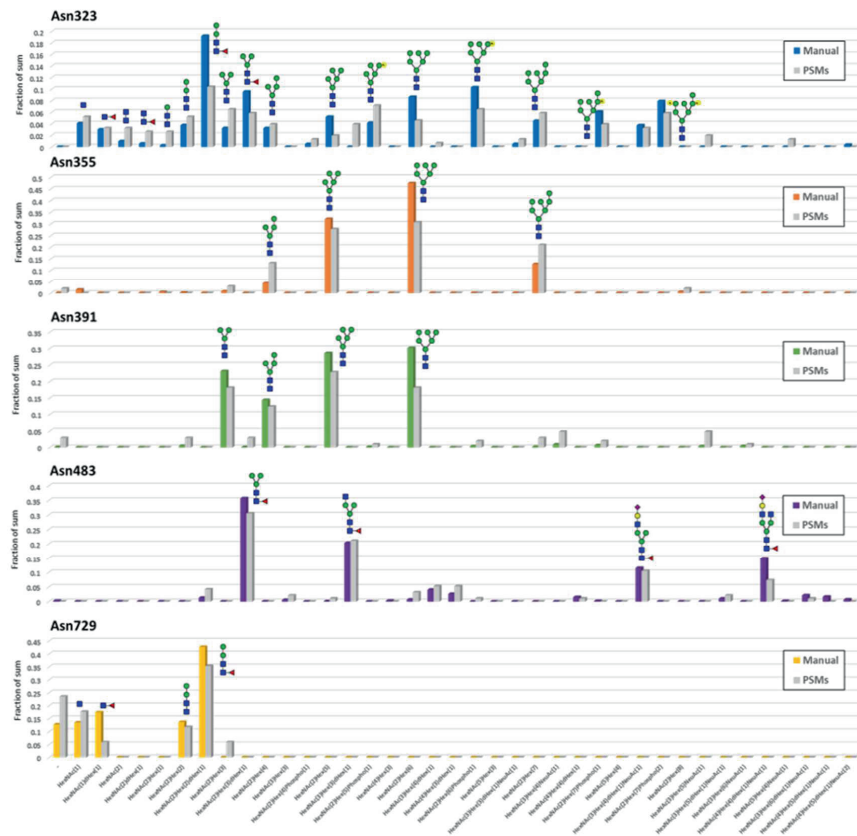


Figure S2: Myeloperoxidase MS¹ integration vs. MS² (PSM) counting.

The following supplementary tables are not printed because they are oversized.

Supplementary Table S1: Proteome specific search and label-free quantification result.

Supplementary Table S2: Proteome non-specific search and label-free quantification result.

Supplementary Table S3: *N*-glycan compositions used for the identification of glycopeptides.

Supplementary Table S4: Identified *N*-glycopeptides in neutrophils.

Chapter 7

Summary and Outlook

Summary

Protein glycosylation is an exceptionally complex PTM due to the high structural diversity of glycans, which are key substances of many fundamental cellular and molecular processes underlying various structural roles across all living organisms. Regarding structural and conformational aspects of glycans, subtle changes in glycan structures can influence the proteins' biological function. Consequently, as described in **Chapter 1**, aberrant glycosylation associates with many diseases such as cancer and various other genetic diseases. Therefore, there is a long-standing interest in developing methods to study the intact glycopeptides and released glycans from complex biological samples. These methods have offered new opportunities to expand our knowledge of protein glycosylation and provided detailed information on the glycan composition. During the four years of my Ph.D., I studied protein glycosylation by using two different mass spectrometry (MS) approaches; native MS and glycopeptide-centric analysis MS. High-resolution native MS has been applied for the in-depth characterization of intact proteins, including the comprehensive evaluation of glycoproteoform heterogeneity and discovery of sequence variants. Peptide-centric MS analysis has been used to complement native MS data and characterize large numbers of glycoproteins in complex biological mixtures. To this end, glycopeptides have been enriched by a novel high-throughput method prior to the MS analysis in order to reduce sample complexity and increase glycopeptide sensitivity.

The in plasma abundant present protein fetuin was discovered already 80 years ago, and has been widely used as a standard glycoprotein for method development in the glycoproteomics field. Despite years of research, however, the structure and biological function of fetuin remain still somewhat unclear. In **Chapter 2**, using native MS, I analyzed and compared intact fetuin, originating from human, bovine, and recombinant sources. From the analysis of human fetuin (hFet), I provided for the first time substantial evidence that the structure of hFet comprises of a heavy A chain with a missing C-terminal arginine and a light B chain. These two chains are connected by a single disulfide bridge. Further, the occupancy of one out of two phosphorylation sites occurring on hFet was shown to be 100%. Although the site-specific glycosylation pattern of hFet has been well-documented, native MS analysis qualitatively and quantitatively revealed overall glycosylation microheterogeneity of hFet and provided some new surprising findings. When comparing hFet with other two fetuins produced by the same gene but from different biological sources, significant differences were observed in their primary structures, PTMs including N-/ O-glycosylation and phosphorylation patterns. One important take-home message from this study was that described structural differences between plasma derived hFet and recombinant rhFet should be carefully considered in any future functional study utilizing rhFet.

After I had demonstrated how the same gene could eventually produce distinct protein products, I dedicated my next research project to the question of whether hFet was also different among

individuals. To address this question, in **Chapter 3**, I purified endogenous hFet from serum of several individual donors, including healthy and septic donors, using ion-exchange chromatography. Subsequently, I characterized in detail proteoform profiles of the purified hFet using native MS combining with peptide-centric MS. The native spectra revealed that hFet isolated from these individuals were distinct due to gene polymorphism and different glycosylation occupancies of hFet. The human serum hFet is a highly genetically polymorphic gene, resulting in three genotypes AHSG*1, AHSG*2, and heterozygous AHSG 1/2. The difference in the backbone amino acid sequences between AHSG*1 and AHSG*2 are caused by a double mutation, involving one known O-glycosylation site. The most intriguing finding in this work was that this O-glycosylation site from AHSG*1 was fully modified by O-glycans, whereas this site from AHSG*2 was mostly unoccupied, indicating a drastic effect of the gene polymorphism on the proteins' glycosylation pattern. Furthermore, an algorithm was developed in this work, allowing to semi-automatically and site-specifically annotate the glycan combination of each proteoform in the native mass spectra. hFet is an acute-phase protein, which may change its glycosylation pattern during inflammatory processes. Our results in this work also showed significantly increased fucosylation levels of hFet proteoforms isolated from septic patients compared to healthy people.

In Chapter 4, we performed a longitudinal study in a plasma sample cohort, containing ten individual septic patients across four time-points (T1-T4). The original goal in this work was the exploration of potential glycoprotein biomarkers for sepsis. Label-free quantitative proteomics experiment on plasma samples was first performed to determine changes in patient's plasma proteomes over time. Our analysis revealed around ten upregulated proteins in both T2 and T3 compared to T1, including some acute-phase proteins such as C-reactive protein (CRP) and AACT. Abundance levels of these proteins reverted back to their initial (T1) values at T4. AACT harboring six N-glycosylation sites exhibits high glycosylation heterogeneity. Therefore, we decided to further purify it using IEX from an individual's plasma and measure it on native MS to acquire a comprehensive picture of its glycosylation pattern, and to investigate whether the changes in glycosylation correlate with the four time points. When we compared the glycoproteoform profiles of AACT from T1 with T2 and T3, we observed a slight shift in molecular mass distribution towards higher masses caused by an increased level of fucosylation, and glycan branching. Surprisingly, this remodeling continued even in T4, where we reproducibly observed an even greater extent of glycan remodeling, suggesting the changes in glycosylation did not follow the same trend as the corresponding protein abundances obtained by bottom-up proteomics analysis. This remarkable observation indicates that the glycan micro-heterogeneity of AACT reflects some so far neglected physiological processes, which take place after the levels

of acute-phase proteins drop back to the levels before sepsis occurred, in a time when patients are already released from the ICU.

To gain a better understanding of the biological effects in the changes of glycan heterogeneity at a given site, a broad and deep study of the glycoproteome from a biological sample is needed. Human milk, a rich source of nutrition containing a variety of bioactive components, attracted my attention because I noticed that most of the glycoproteome studies on milk had been done solely on the level of released glycans, and only a few studies assessed heterogeneity with site-specific information. Besides, very little quantitative information on site-specific glycosylation of milk proteins during lactation is known. The difficulties of human milk N-glycopeptide characterization arise from the broad dynamic range of its protein concentrations and the enormous complexity of its protein glycosylation. In **Chapter 5**, an automated HILIC-based platform was developed to perform an in-depth, large-scale, and high-throughput intact N-glycopeptide identification in human milk, taking samples from one donor across nine time points. By using this platform, we identified 43 new glycosylation sites on 32 glycoproteins, including one on the well-studied phosphoprotein α -S1-casein. In addition to this, we developed a targeted method to obtain label-free quantification of 287 glycopeptides from 50 proteins across the lactational period. Changes in degrees of branched and fucosylated N-glycans of these quantified milk proteins were observed, which can contribute to our knowledge of milk biogenesis and benefits to the developing infant.

In my final chapter, I used the above described methods to analyze in depth the glycoproteome of neutrophils. The most interesting observation in these studies was that a subset of neutrophil glycoproteins carry quite a few glycosylations rarely observed in plasma proteins, notably paucimannose glycans and phosphomannose glycans. As neutrophils belong to some of the most important immune cells playing a role in the body's first line of defense against pathogen infection, neutrophils glycoproteins may be functionally associated with the inflammation processes. Therefore, in **Chapter 6**, we applied in parallel two methods with/without HILIC-based glycopeptide enrichment prior to MS analysis to characterize in detail the neutrophil N-glycoproteome. The data revealed the glycan distribution of each glycosylation site for many high abundant glycoproteins originating from the primary (azurophil) granules. Importantly, the majority of paucimannose glycans could only be detected in the method without HILIC enrichment. Overall, our data demonstrated the site-specific glycosylation of more than 200 individual neutrophil glycoproteins and provided an avenue forward to continue facilitating future research in the roles played by neutrophils in inflammation.

Nederlandse samenvatting

Glycosylering is een bijzonder complexe posttranslationale modificatie dankzij de grote structurele diversiteit van glycanen, de sleutelstukken van vele fundamentele cellulaire en moleculaire processen welke een onderliggende rol spelen in alle levende organismen. Als het gaat om de structurele en conformationele aspecten van glycanen kunnen subtiele veranderingen in de glycaanstructuur de biologische functies van eiwitten al beïnvloeden. Derhalve, zoals beschreven in hoofdstuk 1, wordt afwijkende glycosylering geassocieerd met vele ziektes als kanker en andere verscheidene genetische aandoeningen. Hierom is er een reeds lang bestaande interesse in de ontwikkeling van methodes om intacte glycopeptiden of glycanen, losgemaakt uit complexe biologische monsters, te bestuderen. Deze methoden hebben nieuwe mogelijkheden geboden om onze kennis van glycosylering uit te breiden en ons voorzien van gedetailleerde informatie over de compositie van glycanen. In de vier jaren van mijn PhD heb ik glycosylering door middel van twee verschillende massaspectrometrie (MS) methodes bestudeerd; natieve MS en MS-analyse waar glycopeptiden centraal staan (PC MS, (glyco)peptide centric MS). Natieve MS met hoge resolutie is toegepast voor een diepgaande karakterisering van intacte eiwitten, inclusief de uitgebreide evaluatie van glycoproteovorm-heterogeniteit en de ontdekking van sequentievarianten. PC MS is gebruikt om de natieve MS-data te complementeren en om grote aantallen glycoproteïnen in complexe biologische mengsels te karakteriseren. Hiertoe zijn de glycopeptiden verrijkt door een nieuw uitgevonden methode met hoge doorvoer voorgaand aan de MS-analyse, om monstercomplexiteit te verminderen en de gevoeligheid voor glycopeptiden te verhogen.

Fetuiene, een in plasma rijk aanwezig eiwit, was al 80 jaar geleden ontdekt en is wijdverspreid gebruikt als een standaard glycoproteïne ten behoeve van de ontwikkeling van methodes op het gebied van glycoproteomica. Echter blijven, ondanks jaren van onderzoek, de structuur en biologische functie van fetuine enigszins onduidelijk. In hoofdstuk 2 analyseer en vergelijk ik, met gebruik van natieve MS, intacte fetuïne afkomstig van menselijke, runderen en recombinante bronnen. Vanuit de analyse van menselijk fetuïne (hFet) bied ik voor het eerst substantieel bewijs dat de structuur van hFet is opgebouwd uit een zware A-keten met een missende C-terminus arginine en een lichte B-keten. Deze twee ketens zijn met elkaar verbonden via een enkele disulfidebrug. Aanvullend is aangetoond dat de bezetting van één van de twee fosforyleringsliggingen welke voorkomen op hFet 100% bezet is. Hoewel het liggings-specifieke glycosyleringspatroon van hFet goed beschreven is, onthulde natieve MS-analyse zowel op kwalitatieve als kwantitatieve wijze de globale micro-heterogene glycosylering van hFet en leverde enkele nieuwe, verrassende bevindingen. Toen hFet werd vergeleken met twee andere fetuïnen geproduceerd door hetzelfde gen maar afkomstig van verschillende biologische bronnen, konden significante verschillen in hun primaire structuur worden geobserveerd; posttranslationale

modificaties inclusief N- en O-glycosylering en fosforyleringspatronen. Een belangrijke boodschap van deze studie was dat de beschreven structurele verschillen tussen uit plasma ontleend hFet en recombinant rhFet zorgvuldig moeten worden overwogen in toekomstige functionele studies waarin rhFet wordt gebruikt.

Nadat ik had gedemonstreerd hoe eenzelfde gen uiteindelijk aparte eiwitproducten kon produceren, wijdde ik mijn volgende onderzoeksproject aan de vraag of hFet ook verschillend was in individuen. Om deze vraag in hoofdstuk 8 te benaderen zuiverde ik endogeen hFet van enkele individuele donoren op, inclusief dat van gezonde en in-sepsis donoren, met gebruik van een ionenuitwisselingschromatografie (IEX, ion exchange chromatography). Daaropvolgend karakteriseerde ik gedetailleerd de proteovorm profielen van het opgezuiverde hFet door middel van natieve MS in combinatie met PC MS. De natieve spectra onthulden dat hFet geïsoleerd van deze individuen verschillend was dankzij genetische polymorfismen en afwijkende glycosyleringsbezettingen van hFet. Het menselijk serum hFet is een zeer genetisch polymorfisch gen wat resulteert in drie genotypen; AHSG*1, AHSG*2 en het heterozygote AHSG1/2. Het verschil in de aminozuursequenties van de ruggengraat tussen AHSG*1 en AHSG*2 wordt veroorzaakt door een dubbele mutatie waarbij één bekende O-glycosyleringsligging is betrokken. De meest intrigerende bevinding van deze studie was dat deze O-glycosyleringsligging van AHSG*1 volledig gemodificeerd was door O-glycanen, waarbij deze ligging in AHSG*2 juist voornamelijk onbezet was, wat een indicatie is van een drastisch effect van genpolymorfisme op de glycosyleringspatronen van het eiwit. Verder is bij deze studie een algoritme ontwikkeld wat de mogelijkheid geeft om semiautomatisch en liggings-specifiek de glycaancombinatie van elke proteovorm in natieve massaspectra te benoemen. hFet is een acutefase-eiwit, wat zijn glycosyleringspatroon kan veranderen tijdens ontstekingsprocessen. Onze resultaten van dit werk toonden ook significantie toenames in fucosyleringsniveaus van hFet proteovormen geïsoleerd van sepsis-patiënten, in vergelijking met gezonde personen.

In hoofdstuk 4 hebben we een longitudinaal onderzoek uitgevoerd met een plasma monster cohort, bestaande uit tien individuele sepsis-patiënten gespreid over vier tijdstippen (T1-T4). Het oorspronkelijke doel van deze studie was om potentiële bio markers voor sepsis te verkennen. Eerst was een label-vrij kwantitatief proteomica experiment gedaan op plasma monsters om veranderingen in de proteomen van de patiënten hun plasma vast te stellen over een periode van tijd. Onze analyse onthulde rond de tien in waarde verhoogde eiwitten in zowel T2 als T3, vergeleken met T1, inclusief enkele acutefase eiwitten zoals C-reactief eiwit (CRP) en AACT. De hoeveelheden van deze eiwitten keerden terug naar de beginwaardes (van T1) op T4. AACT, met zes N-glycosylering plekken, vertoonde een hoge glycosylering-heterogeniteit. Hierdoor besloten we het verder op te zuiveren uit een individu's plasma met gebruik van IEX en het te meten met natieve MS om een uitgebreid beeld van zijn glycosyleringspatroon te vergaren, en te

onderzoeken of de veranderingen in glycosylering overeenkomen met de vier tijdstippen. Wanneer we de glycoproteovorm profielen van AACT van T1 vergeleken met T2 en T3 observeerden we een kleine verschuiving van de moleculaire massadistributie naar hogere massa's, veroorzaakt door een verhoogd niveau van fucosylering en glycaan-vertakking. Verassend genoeg bleef deze hermodellering doorgaan op T4, waar we een reproduceerbaar maal een nog grotere glycaan hermodellering observeerden, wat suggereert dat de veranderingen in glycosylering niet dezelfde trend volgens als de overeenkomende eiwit-niveaus die werden gemeten met PC MS. Deze opmerkelijke observatie geeft aan dat de microheterogeniteit van de glycanen van AACT een fysiologisch proces reflecteert, wat tot nu toe links is laten liggen, en plaats vindt nadat de niveaus van acutefase-eiwitten teruggezak zijn naar de niveaus van voor dat sepsis plaatsvond, in de periode dat patiënten al ontslagen zijn van de ICU.

Om een beter beeld te krijgen van de biologische effecten van de veranderingen van glycaan heterogeniteit op een specifieke ligging, is een brede en diepe studie van het glycoproteoom van een biologisch monster noodzakelijk. Menselijke melk, een rijke bron van voedingsstoffen bestaande uit een variëteit van bioactieve componenten, trok mijn aandacht omdat ik opmerkte dat de meeste glycoproteoom-studies naar melk enkel waren gedaan op het gebied van losgemaakt glycanen, en slechts een paar studies beoordeelden de heterogeniteit met liggingspecifieke informatie. Daarnaast is er zeer weinig kwantitatieve informatie bekend over de liggingspecifieke glycosylering van melkeiwitten tijdens het geven van borstvoeding. De moeilijke aspecten van karakterisering van N-glycopeptiden van menselijke melk komen voort uit het brede 'dynamische bereik' van de eiwit concentraties en de enorme complexiteit van de eiwitglycosylering. In hoofdstuk 5 is een geautomatiseerd platform op basis van HILIC ontwikkeld om een diepgaande, grootschalige en hoge-doorvoer identificatie van intacte N-glycopeptiden in menselijke melk uit te voeren, door monsters van een donor te nemen in een periode van negen tijdstippen. Door dit platform te gebruiken, identificeerden we 43 nieuwe glycosyleringsliggingen en 32 glyco-eiwitten, waaronder het goed beschreven alfa-S1-caseïne. Hieraan toevoegend ontwikkelden we een doelmatige methode om labelvrij 287 glycopeptiden te kwantificeren, afkomstig van 50 eiwitten uit de periode van borstvoeding. Veranderingen in de gradaties van vertakte en gefucosyleerde N-glycanen van deze gekwantificeerde melkeiwitten werden geobserveerd, wat kan bijdragen aan onze kennis van de biologische wording van melk en de voordelen voor het ontwikkelende kind.

In mijn laatste hoofdstuk gebruikte ik de hierboven beschreven methodes om het glycoproteoom van neutrofielen in diepte te analyseren. De meest interessante observatie van deze onderzoeken was dat een subset van neutrofiel glycoproteïnen nogal wat glycosyleringen draagt welke zelden worden geobserveerd in plasmaeiwitten, in het bijzonder paucimannose glycanen en fosfomannose glycanen. Aangezien neutrofielen behoren tot een van de meest belangrijke

soorten immuun cellen die een rol spelen in de eerste verdedigingslinie van het lichaam tegen infecties van pathogenen, zouden deze neutrofiel-glycoproteïnen kunnen worden geassocieerd met het ontstekingsproces. Daarom pasten wij in hoofdstuk 6 in parallel twee methoden toe, met of zonder HILIC-gebaseerde verrijking van glycopeptiden voorgaand aan MS-analyse, om het neutrofiel N-glycoproteoom in detail te karakteriseren. De data onthulde dat de glycaandistributie van elke glycosyleringsligging voor veel van de abundante glycoproteïnen voorkomt uit de primaire (azurofiele) granulen. Hierbij is belangrijk dat de paucimannose glycanen alleen konden worden gedetecteerd met behulp van de methode zonder HILIC-verrijking. Onze data demonstreert globaal de liggingsspecifieke glycosylering van meer dan 200 individuele neutrofiel glycoeiwitten en voorziet van een pad naar de vergemakkelijking van toekomstig onderzoek en de rol die wordt gespeeld door neutrofielen bij een ontsteking.

Outlook

Analysis of protein glycosylation by using MS-based methods has relied until recently primarily on targeting either the released glycans or the intact glycopeptides. In the last years, combining native MS with various other MS-based approaches, such as bottom-up LC-MS and top-down LC-MS, has substantially advanced the more in-depth characterization of all discrete glycoforms of a protein. Systematic integration of state-of-the-art analytical methods with hybrid MS approaches may generate data at the glycan, glycopeptides, and glycoproteins levels, which significantly improve our understanding of the unique nature of glycoproteins. Moreover, these analytical methods can provide information about higher-order structures, unknown PTMs, mutations, and sequence variants. In the following paragraphs, I will discuss some potential future directions of the research based on the results described in my thesis in Chapters 2-6. Further, I will briefly describe the importance of considering glycosylation when developing a vaccine against a virus.

Monitoring allele-specific proteoform profiles in a heterozygote donor

As described in Chapter 3, native MS measurement can differentiate genetic variants of proteins based on the observed mass differences between each genetic variant in the MS spectrum. Therefore, it is possible to monitor allele-specific protein expression directly at the intact protein level, if the genetic variants produce resolvable mass shifts. Although genomic analysis has shown numerous different alleles for many polymorphic human proteins, the detailed molecular basis and differential functioning of these proteins remains largely unknown. Comprehensive proteoform profiles of intact proteins contain quantifiable information of all molecular species present, including those carrying the mutations. As an example of such an analysis, I show here some preliminary data on α -1-antitrypsin (AAT). Human plasma AAT functions as a protease inhibitor and is produced mainly by liver cells. AAT deficiency (AATD) is the only proven genetic risk factor associated with the development of chronic obstructive pulmonary disease (COPD)¹. AAT is encoded by the SERPINA1 gene, which contains quite a few sequence variants, originating from approximately 100 identified alleles². Among these alleles, four common variants (M1, M2, M3, and M4) are frequently distributed in healthy people. Two rare variants, Z (Glu342Lys) and S (Glu264Val), are the most deleterious mutations, leading to the severe AATD³. **Figure 1** displays preliminary data from native MS measurements of AAT, which are represented by the proteoform profiles of AAT purified directly from the plasma of five individual donors. Four individuals out of five carry different genotypes, including two heterozygotes (Donor 1, 2 and 5) and two homozygotes (Donor 3 and 4). These mutations in the backbone mass sequence affect the molecular mass of the protein, introducing a mass shift

between the corresponding proteoform signals. Notably, the native MS spectra of AAT from donors of the heterozygous genotype (**Figure 1 a, b and e**) show more complex proteoform profiles compared to the spectra from homozygous genotype donors (**Figure 1 c, d**). The proteoform profiles of heterozygotes donors represent a combination of all proteoforms generated by M1A allele and M1V allele. The detected mass difference of 28.03 Da between the corresponding peaks generated from M1A and M1V is induced by a mass difference between the amino acids Ala and Val in the mutation site. Interestingly, in heterozygote donors, the data revealed donor and allele-specific protein abundances. The relative abundances of corresponding proteoforms generated by M1A allele and M1V allele are distinct (**Figure 1 a, b**). In donor 1, proteoforms generated by M1A allele and M1V allele are equally abundant (M1A:M1V is around 0.99:1) (**Figure 1a**), whereas in donor 2 the predominant proteoforms originate from M1A (M1A:M1V is around 5:1)(**Figure 1b**). This intriguing result can have many causes; it either implies that the M1A and M1V alleles are not equally expressed in the liver of these individuals, but it may also mean that they are not equally secreted into the blood stream, and also it may be that their lifetime/stability in the bloodstream is different. Thus, these observations require further validation, whereby it would be interesting to investigate whether these differences in abundance in heterozygote donors are regulated at the transcriptional or post-transcriptional level. The next experiments should include an extension of the analysis to more donors, especially heterozygotes. Also, characterizing AAT derived from Z and S genotypes and comparing their proteoform profiles AAT derived from other common genotypes could bring some understanding to diseases related to the AAT deficiency.

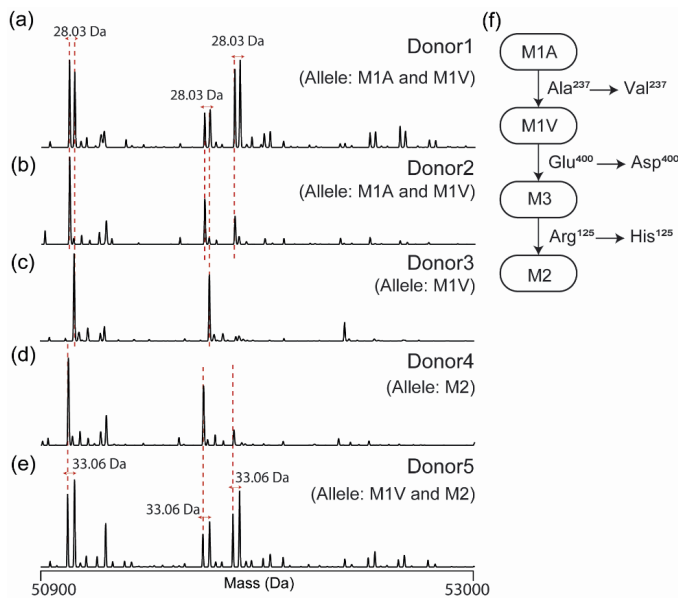


Figure 1. Different genotypes of AAT lead to different proteoform profiles.

Zero-charge deconvoluted native mass spectra of AAT extracted from the plasma of five individual donors, including 3 exhibiting a heterozygous genotype: (a) Donor 1 and (b) Donor 2 carry both the M1A and M1V allele; (c) Donor 5 carries both the M1V and M2 allele, and two homozygous genotypes: (d) Donor 3 carries the M1V allele and (e) Donor 4 carries the M2 allele. Sequence differences among the major AAT variants are depicted in (f). The sequence of AAT from M1V differs from M1A by one amino acid from alanine to valine at position 237, which introduces the mass difference of 28.03 Da. The sequence of AAT from M2 differs from M1V by two amino acid from glutamic acid to asparagine and arginine to histidine at position 400 and 125, respectively. This introduces the mass difference of 33.06 Da³.

Investigating the role of glycosylation in protein-protein and protein-drug interactions

Native MS largely retains ionized protein molecules in their folded state, which allows for the analysis of non-covalent interactions. The ability to preserve protein-protein interactions or protein-drug interactions can be utilized for quantitative analysis of proteoform profiles of these assemblies and investigation of the role of glycosylation in their interactions. Acute phase reactants are a great example of heavily glycosylated proteins involved in many protein-protein interactions. So far it is largely unknown what the specific role of protein glycosylation is in these interactions. In the following part, I will suggest some future research related to AACT, which was described in detail in Chapter 4. AACT is one of the serpin superfamily proteins, which mainly function as physiological inhibitors of the serine proteases of neutrophil azurophil granules such as elastase, proteinase-3, and CathG. These proteases play an essential role in the antimicrobial defense. However, when they are released in an uncontrolled manner during phagocytosis, other matrix proteins and immune defense molecules can also become degraded, thus leading to coagulation disorders, immunodeficiency, or vascular occlusion⁴. It has been known that AACT is capable of forming a denaturant-stable complex with CathG via an irreversible two-step mechanism involving an intermediate complex⁴, which can stabilize AACT-CathG complex. Both AACT and CathG are N-glycoproteins, presenting highly flexible and heterogeneous N-glycans on their surfaces. These N-glycans are likely to regulate protease-inhibitor interaction, but the detailed impact of microheterogeneity on the interaction remains unknown. In similar fashion as in Di Wu et.al⁵, a series of affinity binding experiments can be performed to explore the influence of different glycoproteoforms on AACT-CathG interaction. CathG should then be immobilized to an agarose resin (Cath-agarose) and subsequently incubated with AACT. Such assay would allow for an in-depth quantitative analysis of proteoform profiles of AACT with different affinities toward CathG by high-resolution native MS. The acquired data could reveal relationships between different AACT affinity properties toward CathG, and the role of subtle changes in glycosylation profiles. Also, it may be possible in the near future to design inhibitors with functional glycan structures, which perform similar binding affinities as their natural counterparts.

Glycosylation is important in the design of future vaccines

During the time of writing my Ph.D. thesis, the world has been overwhelmed by the tremendous crisis caused by the pandemic spread of the SARS-CoV-2 virus disease (Covid-19) which triggered intense global research activity to develop a vaccine against this virus. Currently, most vaccines use proteins of the pathogens to elicit antibodies, which enable them to neutralize viral pathogens and prevent further infection of host cells⁶. However, the effectiveness of many existing vaccines is quite low⁷, mainly due to the following reasons. First, immunodominant regions of virus antigens are extremely variable in sequence due to the highly frequent mutations of amino acid residues, thereby hampering the elicitation for broader immune protection. Second, the surface of both host cells and viral spike proteins is surrounded by complex glycans, which is capable of influencing interactions between the viruses and host organisms that control viral pathogenesis⁸. As described in Chapter 1-2, changes in glycosylation are related to host-pathogen interactions, which can help pathogens to evade host antibody and innate immune defenses. For example, by recombinantly adding more glycosylation sites onto the influenza hemagglutinin (HA) protein, the immune response in infected mice vaccinated by mutated virus strain can not be effectively induced against the original virus strain⁹. In order to improve the efficacy of vaccines, it is necessary to characterize the glycosylation of both host cell proteins and the viral spike proteins and understand the full glycosylation effects on immunodominance. Considering glycosylation effects on immunodominance in a paradigm level of future vaccine design, more broadly neutralizing vaccines can be developed and protect us away from the devastating diseases such as Covid-19.

References

- (1) Craig, T. J.; Paula, M. Advances in Managing COPD Related to $\alpha 1$ - Antitrypsin Deficiency : An under-Recognized Genetic Disorder. *Allergy* 2018, 73 (11), 2110–2121.
- (2) Al-Jameil, N.; Hassan, A. A.; Buhairan, A.; Hassanato, R.; Isac, S. R.; Al-Otaiby, M.; Al-Maarik, B.; Al-Ajeyan, I. Genotyping Diagnosis of Alpha-1 Antitrypsin Deficiency in Saudi Adults with Liver Cirrhosis. *Med. (United States)* 2017, 96 (6), e6071.
- (3) Brantly, M.; Nukiwa, T.; Crystal, R. G. Molecular Basis of Alpha-1 -Antitrypsin Deficiency. *Am. J. Med.* 1988, 84, 13–31.
- (4) Duranton, J.; Adam, C.; Bieth, J. G. Kinetic Mechanism of the Inhibition of Cathepsin G by A1- Antichymotrypsin and A1-Proteinase Inhibitor. *Biochemistry* 1998, 37 (32), 11239–11245.
- (5) Wu, D.; Struwe, W. B.; Harvey, D. J.; Ferguson, M. A. J.; Robinson, C. V. N-Glycan Microheterogeneity Regulates Interactions of Plasma Proteins. *Proc. Natl. Acad. Sci. U. S. A.* 2018, 115 (35), 8763–8768.
- (6) Kwong, P. D.; Mascola, J. R.; Nabel, G. J. Rational Design of Vaccines to Elicit Broadly Neutralizing Antibodies to HIV-1. *Cold Spring Harb. Perspect. Med.* 2011, 1 (1).
- (7) States, U.; Dawood, F. S.; Chung, J. R.; Kim, S. S.; Zimmerman, R. K.; Nowalk, M. P. Interim Estimates of 2019–20 Seasonal Influenza Vaccine Effectiveness — United States, February 2020. *Weekly* 2020, 69 (7), 177–182.
- (8) Tate, M. D.; Job, E. R.; Deng, Y. M.; Gunalan, V.; Maurer-Stroh, S.; Reading, P. C. Playing Hide and Seek: How Glycosylation of the Influenza Virus Hemagglutinin Can Modulate the Immune Response to Infection. *Viruses* 2014, 6 (3), 1294–1316.
- (9) Sun, X.; Jayaraman, A.; ManiPrasad, P.; Raman, R.; Houser, K. V.; Pappas, C.; Zeng, H.; Sasisekharan, R.; Katz, J. M.; Tumpey, T. M. N-Linked Glycosylation of the Hemagglutinin Protein Influences Virulence and Antigenicity of the 1918 Pandemic and Seasonal H1N1 Influenza A Viruses. *J. Virol.* 2013, 87 (15), 8756–8766.

Appendix

Curriculum Vitae

List of publications

Acknowledgments

Curriculum Vitae

Yu-Hsien Lin was born on 5th December 1990 in Taichung, Taiwan. After obtaining the Bachelor of Science of Chemistry at National Taiwan Normal University in Taipei, Taiwan in 2014, he subsequently decided to pursue his master study in chemistry in the same university. He enrolled in the mass spectrometry and proteomics group led by Prof. Dr. Yu-Ju Chen at the Chemistry department, Academia Sinica research institute in Taipei, Taiwan. During this period he applied mass spectrometry to analyze post-translational modifications modified on proteins. He performed his master's thesis project in the method development of glycopeptide enrichment and ultimately obtained his Master of Science of Chemistry in 2016.

In the same year, he relocated to the Netherlands and enrolled in the Biomolecular mass spectrometry and proteomics group under the supervision of Prof. Dr. A. J. R. Heck and his daily supervisors Dr. Vojtech Franc and Dr. Karli R. Reiding. In Utrecht, his research project primarily focused on applying hybrid mass spectrometry approach, which combined native mass spectrometry and peptide-centric mass spectrometry for characterization of human plasma glycoproteins.

List of publications

1. **Yu-Hsien Lin**, **Jing Zhu**, Dingess, Kelly; Mank, Marko; Stahl, Bernd; Heck, Albert. A quantitative longitudinal inventory of the N-glycoproteome of human milk from a single donor reveals the highly variable repertoire and dynamic site-specific changes. *J Proteome Res.* 2020, 19 (5), 1941-1952.
2. **Yu-Hsien Lin**, Jing Zhu, Sander Meijer, Vojtech Franc, Albert J. R. Heck. Glycoproteogenomics: A Frequent Gene Polymorphism Affects the Glycosylation Pattern of the Human Serum Fetuin/ α -2-HS-Glycoprotein. *Mol. Cell. Proteomics.* 2019, 18 (8), 1479-1490.
3. **Yu-Hsien Lin**, Vojtech Franc, Albert J. R. Heck. Similar Albeit Not the Same: In-Depth Analysis of Proteoforms of Human Serum, Bovine Serum, and Recombinant Human Fetuin. *J. Proteome Res.* 17 (8), 2861-2869.

Acknowledgments

Four years have passed since I joined the fantastic “Heck group” in 2016. On this long PhD journey which was full of challenges and obstacles, I would not have been able to arrive at my destination without all the people that supported me along the way. I am beyond grateful to everyone who has been with me during these years. Everything we have been through together has created memories that will last a long time.

Albert, I express my warm thanks to you for your guidance during my PhD journey. In 2015, you gave an inspiring talk in the Academia Sinica, Taiwan. After that presentation I spoke to you and asked for an opportunity to join your research group. We had a short discussion about what I had done for my Master's projects. My first impression to you was as an enthusiastic, passionate, creative and disciplined scientist. The most important thing that I learned from you is to “never ignore any small findings during research”. I remember every time when I discussed data that I felt bored with, you always said rather that it was interesting. With your encouragement, I kept a positive outlook on my projects and had more confidence. I enjoyed the time working with you during these years.

Vojtech, you were not only my daily supervisor but also my best friend in the group. You helped me a lot during these past several years. Colleagues were always joking that I was your son but instead, I think you were my brother. You were highly organized and attentive during our work, and showed me what are the essential elements for being a research scientist. We had lunch together almost every day. You were talkative and shared many funny stories with me. I could not imagine my life without you in these years. Thank you for being there for me. I definitely will visit you during my holidays.

Karli, you became my daily supervisor during my last year. You had many creative insights, which often gave me ideas about what to do with my data. You gave me valuable lessons about the basic knowledge of glycans. You told me you love the feeling of sharing your knowledge to others and you have the passion for science. I believe you will become an amazing professor in the future.

I still remember the first day I came into the “Heck group”. I felt nervous to be joining such a big group of nearly 50 people from all over the world. However, people in the group were so friendly and easygoing and I was grateful to join this international group. **Tomi**, you came to me during my first day and told me “No one does glyco stuff in this lab except me. So you should start thinking about what you can do in this lab and if you have any question, you can just come to me.” I felt impressed by your kind words on my first day. You were such a “walking library” about glycoscience. I was happy to be your officemate in my last year and enjoyed playing table tennis with you. I believe you will become the most famous glycoscientist in the world in the near future!

Miao, thank you for helping me out when I arrived in the Netherlands. With your assistance, I quickly became involved in the group. I enjoyed my time playing badminton with you and **Tony**. Now you have become a PI and have your own group. I can't wait to see you conduct more interesting research and share that knowledge with us. **Jing**, I know you were always annoyed by the many stupid questions I asked you. However, you still patiently answered all my questions. Within the three years of my time in the lab, we worked together and published two journal articles. You also explained to me how to use **R** to do data analysis and showed me some tricks during native MS measurements. Furthermore, you often shared your opinions and advice for the design of my experiments, although these were not your projects. Congratulations for your promotion in China and wish you all the best in your next steps in the field of nutrition. **Yang**, thank you for guiding me to take over your work. You also gave me a lot of useful advice, which prepared me to go through the job interview at your company. **Anja**, thanks for giving me a first lesson of native MS. **Dominique**, I started learning how to fix the LC system with you. I truly appreciate your patience for giving me instructions. I wish you all the best in your new position. **Franziska** and **Laura**, thanks for being operators of the Lumos with me. **Harm, Arjan, Mirjam, Celia, Corine, Henk and Geert**, thank you for taking care of everything in the lab such as maintaining all the instruments, filling empty bottles of buffers, organizing social activities and managing of the server in our lab. **Celine and Esther**, you were my great officemates. You kindly shared lots about Dutch culture with me. It was so nice to have you in my office. Also, thanks to all the borrel and party enthusiasts, **Micheal, Philip, Domenico, Julia, Dario, Barbara, Nadine, Charlotte, Max, Kelly D, Pascal, Maurits, Bastiaan, Johannes, Gardi** and everyone else..

Many thanks to all the members of the lunch group. I was happy to that our group came together and that we could eat warm food every day at the UMC. **Wei**, I was grateful to you for encouraging me and giving me more confidence. **Sem**, during working hours you were so focused and concentrated. However, when we were out of the office, you were always the one who organized many social activities such as drinking parties, playing poker and having dinners together. **Oleg**, I admire that you had a clear goal for your future and kept pursuing it. **Linsey**, you are talented in speaking Chinese! "ni hao"! **Tobias**, you were my best partner for playing tennis. I enjoyed the time that we brainstormed how to improve our skills and indeed we both became better players!

Life would be so boring if I only worked in the lab. I would like to thank all my Chinese friends around me during these years. **大梅、楊敏、胡潔、丁彥、亮姊、張浩**, we went hiking, had BBQs in the park and cooked Chinese food together. It was always a lot of fun with you guys. Also, I would like to thank my Taiwanese friends **飛月和瓊如**. Although we rarely met with each other, it was great fun to talk with you about life in the Netherlands and how we missed Taiwanese food. I would like to also thank my previous supervisor during my Master's study Prof. **Yu-Ju Chen**. You brought me into the field of mass spectrometry and proteomics when I was in

my fresh year as an undergraduate student.

Finally, I would like to thank my family for supporting me during these years. 親愛的爸、媽 還有妹妹，感謝你們讓我無後顧之憂的到一個離開家裡的麼遠的國度長達四年。雖然我一年還是有回家一次，你們也來歐洲玩了兩次，但是在這四年我們相處的時間真的不夠。謝謝你們的體諒，我也順利拿到學位準時畢業了，應該沒讓你們失望吧！親愛的蘇醒，很幸運能在剛到荷蘭沒多久就認識你，謝謝妳陪我走過這些年。就如同妳說的，我們真的平均每三個月就會出去旅遊一次，我們一起在歐洲許多地方留下足跡。四年來我們一起度過每一個平凡的周末，有妳在身邊的日子總是這麼的輕鬆自在。雖然接下來我們要暫時分隔兩地，我相信我們一定能克服這樣的距離和時差繼續朝著我們理想中的未來前進！