

In-Depth *N*-glycoproteomic Analysis of Human Blood Plasma Proteins

Dissertation

zur Erlangung des akademischen Grades

Doktoringenieurin

(Dr.-Ing.)

von M. Sc. Frania Jaqueline Zúñiga Bañuelos

geboren am 25. 04. 1991 in Tepic (Mexiko)

genehmigt durch die Fakultät für Verfahrens- und
Systemtechnik der Otto-von-Guericke Universität
Magdeburg

Gutachter/innen:

Prof. Dr.-Ing. Udo Reichl
Prof. Dr. Manfred Wuhrer
Dr. rer. nat. Erdmann Rapp

Promotionskolloquium am: 20. 06. 2025

Intentionally blank page

Intentionally blank page

Intentionally blank page

Abstract

One of the most common post-translational modifications of proteins is glycosylation. This modification enzymatically attaches a sugar chain, called glycan, to a protein. Glycosylated proteins represent around 45% of the proteins in the eukaryotic domain and about 60% of the proteins secreted to the human blood plasma. Regarding their biosynthesis, glycoproteins have a genetically determined component (the protein) and a variable component (the glycan). The biosynthesis of the latter is highly influenced by biological and environmental factors. There are three protein glycosylation types named *N*-, *O*- and *C*-glycosylation, of which *N*-glycosylation is the most frequent type. *N*-linked glycans are more extended than *O*- and *C*-glycans, can show various compositions, and are attached to a conserved sequon (Asn-X-Ser/Thr, where X is any amino acid except Pro). As a result, the structural diversity of *N*-glycans is extraordinary and the number of *N*-glycosylation sites per protein is variable. Thus, protein *N*-glycosylation increases significantly the proteoform heterogeneity of a biological system.

While the analysis of individual *N*-glycans generates meaningful information, only the site-specific analysis of *N*-glycans can reveal their micro-heterogeneity per protein *N*-glycosylation site. Further, the characterization of site-specific *N*-glycosylation changes across human traits can provide critical insights, for example about a pathological state. Mass spectrometry (MS) is the preferred tool to achieve the site-specific elucidation of *N*-glycans via bottom-up glycoproteomics. The two typical approaches for bottom-up glycoproteomics are matrix-assisted laser desorption/ionization time-of-flight tandem mass spectrometry (MALDI-TOF-MS/MS) and liquid chromatography with tandem mass spectrometry (LC-MS/MS). While MALDI-TOF-MS/MS is a high-throughput technology, LC-MS/MS is convenient for analysis of complex protein mixtures. The application of an adequate sample preparation method and the data analysis workflow is essential for the correct site-specific elucidation of *N*-glycans of a complex protein sample, such as human blood plasma.

Human blood plasma or serum is a fluid that networks organs and tissues. In-depth exploration of the composition of human blood glycoproteins is important, as the relation between alterations of *N*-glycosylation in health and disease has been demonstrated in many instances, such as autoimmune diseases, cancer, congenital disorders, and neurodegenerative disorders. Several strategies have been developed for the site-specific *N*-glycan characterization of human blood plasma proteins. The primary scope is to use the low-abundant glycoproteome as a source of therapeutic, prognostic, or diagnostic biomolecules for treatment decision-making. However, due to various limitations, analytics struggles to identify the low-abundant human blood plasma *N*-glycoproteome.

In this thesis, a sample preparation workflow was established and applied on a standard human blood plasma sample (pool of normal citrated plasma of at least 20 donors) followed by a newly developed data analysis workflow, to achieve in-depth *N*-glycoproteomic analysis of the low-abundant human blood plasma glycoproteome. The established sample preparation workflow features two fractionation strategies for expanding the detection range: first, high-abundant protein depletion and second, a protein fractionation by molecular weight. Next, protein fractions are digested to obtain (glyco-)peptides, and the glycopeptides enriched to enhance the chance for their identification. Two high-resolution MS methods are applied for data collection, featuring different collision fragmentation energies. The stepped fragmentation energy gives glycan and peptide composition information, while the low fragmentation energy provides additional structural information. This MS-based measurement strategy follows methodologies previously established in this research group. The developed data analysis workflow comprises three phases: validation, integration of structural features, and re-annotation. First, data validation, based on a newly proposed decision tree, evaluates the coherence between the observed evidence and the proposed glycopeptide spectra match (gPSM). Second, the integration of structural features incorporates *N*-glycan information observed in the MS spectra acquired at a low fragmentation energy and enables the revision of incorrect gPSM. Finally, the data analysis focuses on re-annotating incorrect gPSMs.

Application of this sample preparation workflow not only allows the identification of a high number of *N*-glycan compositions per *N*-glycosylation site, but also a significant expansion of the detection range of human blood plasma glycoproteins. While the analysis of untreated blood plasma allowed only the detection of glycoproteins in the range from $1 \cdot 10^6$ to $1 \cdot 10^9$ pg/mL, the lower observed protein concentration is $6 \cdot 10^3$ pg/mL with the newly established sample preparation method. In a few instances even the detection of glycoproteins at a lower abundance, like the identification of *N*-glycopeptides derived from cysteine-rich secretory protein LCCL domain-containing 1 (8.2 pg/mL), was achieved. After data validation, a total of 7,867 gPSMs were classified and condensed to 1,929 *N*-glycopeptides referring to 942 different *N*-glycosylation sites that belong to 805 human glycoproteins. Moreover, the data analysis workflow uncovers frequent pitfalls causing incorrect gPSM assignment, such as the identification of two fucoses instead of one neuraminic acid (NeuAc), which needs attention in further method development. By the integration of MS spectra acquired at low fragmentation energy, rare structures (e.g., *N*-glycans holding sialic acid-*N*-acetylhexosamine linkage on one antenna plus an extended *N*-acetylactosamine at the second antenna), are additionally identified. Other example of incorrectly annotated gPSMs, which were successfully re-annotated using the developed workflow, are *N*-glycopeptides bearing three rare *N*-glycan building blocks (presumably monosaccharides as rare capping sugars). Based on mass accuracy it can be inferred that one of these unknown capping sugars is likely glucuronic acid. However, the two other

rare *N*-glycan building blocks, detected on human prothrombin *N*-glycans attached to the N121 site, are still unknown. The presence of these rare *N*-glycan building blocks has been overlooked in earlier studies. Overall, most of the rare *N*-glycans identified are reported here for the first time.

The developed in-depth *N*-glycoproteomic approach was also applied for the analysis of human serum IgA, a clinically relevant glycoprotein, in order to identify rare *N*-glycans. Several of these rare *N*-glycans were reported by glycomic analyses but overlooked using glycoproteomic methods, which hinders the understanding of their role in IgA *N*-glycosylation. Using the developed strategies, the site-specific identification of sulfated and other rare *N*-glycans, such as *O*-acetylated ones, was performed. Optimization of the data analysis setup was conducted with fractionated IgA, and also analyzing two commercial IgA samples with a simple and universal sample preparation method. A key finding of the optimized IgA glycoproteomic analysis is the identification of seven sulfated and four *O*-acetylated sialylated *N*-glycans, which includes seven *N*-glycans not reported before by glycomic analyses. In total eleven rare IgA *N*-glycans were identified for the first time in a site-specific manner for human serum IgA. Also, this study achieved a micro-heterogeneity analysis of IgA *N*-glycosylation that now includes *O*-acetylated and sulfated *N*-glycans. This analysis reflected that sulfated *N*-glycans are predominantly at the IgA tailpiece site, while *O*-acetylated *N*-glycans are detected ten times less abundant than sulfated *N*-glycans. In addition, the MS data obtained from the two IgA samples were also screened for other modifications using oxonium marker ions observed with the in-depth *N*-glycoproteomic analysis of human blood plasma. This allowed the site-specific identification of rare *N*-glycans bearing hexuronic acid (likely glucuronic acid) as capping attached to contaminating proteins in both commercial IgA samples.

In summary, the developed workflow is suitable for expanding our tools to describe the *N*-glycosylation micro-heterogeneity by the identification of rare *N*-glycans, elucidation of *N*-glycan structural features, and detection of glycoproteins in the middle- to low-concentration range of blood plasma proteins. In the future, key features of the data analysis workflow can be transferred to glycoproteomics software tools to generate a more accurate interpretation of MS-based *N*-glycoproteomic measurements. To broaden the possibilities for glycoproteomics research, the workflow presented can be adapted to better characterize other biological samples, such as glycoproteins isolated from different organisms or other body fluids, such as urine, milk, saliva, or mucosal secretions.

Kurzfassung

Eine der häufigsten posttranslationalen Modifikationen von Proteinen ist die Glykosylierung. Bei dieser Modifikation wird enzymatisch eine Zuckerkette, Glykan genannt, an ein Protein angehängt. Glykosylierte Proteine machen etwa 45 % der Proteine im eukaryotischen Bereich und etwa 60 % der in das menschliche Blutplasma ausgeschiedenen Proteine aus. Was ihre Biosynthese betrifft, so haben Glykoproteine eine genetisch festgelegte Komponente (das Protein) und eine variable Komponente (das Glykan). Die Biosynthese des letzteren wird in hohem Maße durch biologische und Umweltfaktoren beeinflusst. Es gibt drei Arten der Proteinglykosylierung, die *N*-, *O*- und *C*-Glykosylierung, von denen die *N*-Glykosylierung die häufigste ist. *N*-gebundene Glykane sind länger als *O*- und *C*-Glykane, können verschiedene Zusammensetzungen aufweisen und sind an eine konservierte Sequenz (Asn-X-Ser/Thr, wobei X eine beliebige Aminosäure außer Pro ist) gebunden. Infolgedessen ist die strukturelle Vielfalt der *N*-Glykane außergewöhnlich und die Anzahl der *N*-Glykosylierungsstellen pro Protein variabel. Somit erhöht die *N*-Glykosylierung von Proteinen die Heterogenität der Proteoformen in einem biologischen System erheblich.

Während die Analyse einzelner *N*-Glykane aussagekräftige Informationen liefert, kann nur die ortsspezifische Analyse von *N*-Glykanen deren Mikroheterogenität pro Protein-*N*-Glykosylierungsstelle aufdecken. Darüber hinaus kann die Charakterisierung ortsspezifischer *N*-Glykosylierungsveränderungen bei menschlichen Merkmalen entscheidende Erkenntnisse liefern, z. B. über einen pathologischen Zustand. Die Massenspektrometrie (MS) ist das bevorzugte Instrument, um die ortsspezifische Aufklärung von *N*-Glykanen durch Bottom-up-Glykoproteomik zu erreichen. Die beiden typischen Ansätze für die Bottom-up-Glykoproteomik sind die Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Tandem-Massenspektrometrie (MALDI-TOF-MS/MS) und die Flüssigkeitschromatographie mit Tandem-Massenspektrometrie (LC-MS/MS). Während die MALDI-TOF-MS/MS eine Hochdurchsatztechnologie ist, eignet sich die LC-MS/MS für die Analyse komplexer Proteingemische. Die Anwendung einer geeigneten Probenvorbereitungsmethode und des Datenanalyse-Prozesses ist für die korrekte ortsspezifische Aufklärung von *N*-Glykanen einer komplexen Proteinprobe, wie z. B. menschliches Blutplasma, von wesentlicher Bedeutung.

Das menschliche Blutplasma oder Serum ist eine Flüssigkeit, die Organe und Gewebe vernetzt. Eine eingehende Untersuchung der Zusammensetzung menschlicher Blutglykoproteine ist wichtig, da der Zusammenhang zwischen Veränderungen der *N*-Glykosylierung in Gesundheit und Krankheit in vielen Fällen nachgewiesen wurde, z. B. bei Autoimmunerkrankungen, Krebs, angeborenen Störungen und

neurodegenerativen Erkrankungen. Für die ortsspezifische Charakterisierung von *N*-Glykanen in menschlichen Blutplasma-Proteinen wurden mehrere Strategien entwickelt. Das Hauptziel besteht darin, die Minorität des Glykoproteoms als Quelle für therapeutische, prognostische oder diagnostische Biomoleküle zur Entscheidungsfindung bei der Behandlung zu nutzen. Aufgrund verschiedener Einschränkungen hat die Analytik jedoch Schwierigkeiten, das schwach vorhandene *N*-Glykoproteom des menschlichen Blutplasmas zu identifizieren.

In dieser Arbeit wurde ein Arbeitsablauf für die Probenvorbereitung etabliert und auf eine Standard-Humanblutplasmaprobe (Pool aus normalem Zitratplasma von mindestens 20 Spendern) angewandt, gefolgt von einem neu entwickelten Arbeitsablauf für die Datenanalyse, um eine eingehende *N*-Glykoproteomanalyse der Minorität der Glykoproteoms des menschlichen Blutplasmas durchzuführen. Der etablierte Arbeitsablauf zur Probenvorbereitung umfasst zwei Fraktionierungsstrategien zur Erweiterung des Nachweisbereichs: erstens die Abreicherung von Proteinen mit hohem Vorkommen und zweitens eine Proteinfractionierung nach Molekulargewicht. Anschließend werden die Proteinfractionen verdaut, um (Glyko-)Peptide zu erhalten, sowie die Glykopeptide angereichert, um die Chancen für ihre Identifizierung zu erhöhen. Zur Datenerfassung werden zwei hochauflösende MS-Methoden mit unterschiedlichen Kollisionsfragmentierungsenergien eingesetzt. Die gestufte Fragmentierungsenergie liefert Informationen über die Glykan- und Peptidzusammensetzung, während die niedrige Fragmentierungsenergie zusätzliche Strukturinformationen liefert. Diese MS-basierte Messstrategie folgt den in dieser Forschungsgruppe bereits etablierten Methoden. Der entwickelte Datenanalyse-Prozess umfasst drei Phasen: Validierung, Integration von Strukturmerkmalen und Neuannotation. Erstens bewertet die Datenvalidierung auf der Grundlage eines neu vorgeschlagenen Entscheidungsbaums die Kohärenz zwischen dem beobachteten Nachweis und der vorgeschlagenen Glykopeptid-Spektren-Übereinstimmung (gPSM). Zweitens werden durch die Integration von Strukturmerkmalen *N*-Glykan-Informationen aus MS-Spektren mit niedriger Fragmentierungsenergie berücksichtigt, was die Überarbeitung eines falschen gPSM ermöglicht. Schließlich konzentriert sich die Datenanalyse auf die Neuannotation falscher gPSMs.

Die Anwendung dieses Probenvorbereitungs-verfahrens ermöglicht nicht nur die Identifizierung einer hohen Anzahl von *N*-Glykan-Zusammensetzungen pro *N*-Glykosylierungsstelle, sondern auch eine deutliche Erweiterung des Nachweisbereichs von Glykoproteinen aus menschlichem Blutplasma. Während die Analyse von unbehandeltem Blutplasma nur den Nachweis von Glykoproteinen im Bereich von $1 \cdot 10^6$ bis $1 \cdot 10^9$ pg/mL erlaubte, liegt die niedrigere beobachtete Proteinkonzentration mit der neu etablierten Probenvorbereitung bei $6 \cdot 10^3$ pg/mL. In einigen wenigen Fällen gelang sogar der Nachweis von Glykoproteinen mit geringerer Häufigkeit, wie z. B. die Identifizierung von *N*-Glykopeptiden aus dem cysteinreichen sekretorischen Protein LCCL domain-containing 1 (8,2 pg/mL). Nach der Datenvalidierung

wurden insgesamt 7.867 gPSMs klassifiziert und zu 1.929 *N*-Glykopeptiden verdichtet, die sich auf 942 verschiedene *N*-Glykosylierungsstellen beziehen, die zu 805 menschlichen Glykoproteinen gehören. Darüber hinaus deckt der Datenanalyse-Prozess häufige Fallstricke auf, die zu einer falschen gPSM-Zuordnung führen, wie z. B. die Identifizierung von zwei Fucosen anstelle von einem Neuraminsäure (NeuAc), was bei der weiteren Methodenentwicklung beachtet werden muss. Durch die Integration von MS-Spektren, die bei niedriger Fragmentierungsenergie aufgenommen wurden, werden zusätzlich seltene Strukturen (z. B. *N*-Glykane mit Sialinsäure-*N*-Acetylhexosamin-Bindung an einer Antenne und einem verlängerten *N*-Acetylactosamin an der zweiten Antenne) identifiziert. So konnten beispielsweise Glykopeptide mit drei seltenen *N*-Glykan-Bausteinen (als Capping-Zucker) als falsch zugeordnete gPSMs erkannt und erfolgreich neu annotiert werden. Anhand der Massengenauigkeit kann gefolgert werden, dass es sich bei einem dieser unbekannten Capping-Zucker wahrscheinlich um Glucuronsäure handelt. Die beiden anderen seltenen *N*-Glykan-Bausteine, die auf menschlichen Prothrombin-*N*-Glykanen an der N121-Stelle nachgewiesen wurden, sind jedoch noch unbekannt. Insgesamt werden die meisten der seltenen *N*-Glykane, die identifiziert wurden, hier zum ersten Mal beschrieben.

Der entwickelte detaillierte *N*-Glykoproteom-Ansatz wurde auch für die Analyse von Humanserum-IgA angewandt, um seltene *N*-Glykane zu identifizieren. Mehrere dieser seltenen *N*-Glykane wurden durch glykomische Analysen entdeckt, aber mit glykoproteomischen Methoden übersehen, was das Verständnis ihrer Rolle bei der *N*-Glykosylierung von IgA erschwert. Mit den entwickelten Strategien wurde die ortsspezifische Identifizierung von sulfatierten und anderen seltenen *N*-Glykanen, wie z. B. *O*-acetylierten, durchgeführt. Die Optimierung der Datenanalyse erfolgte mit fraktioniertem IgA und der Analyse von zwei kommerziellen IgA-Proben mit einer einfachen und universellen Probenvorbereitungsmethode. Ein zentrales Ergebnis der optimierten IgA-Glykoproteomanalyse ist die Identifizierung von sieben sulfatierten und vier *O*-acetylierten sialylierten *N*-Glykanen, darunter sieben *N*-Glykane, die bisher noch nicht durch glykomische Analysen erfasst wurden. Insgesamt wurden elf seltene IgA-*N*-Glykane zum ersten Mal ortsspezifisch für menschliches Serum-IgA identifiziert. Außerdem wurde in dieser Studie eine Mikroheterogenitätsanalyse der IgA-*N*-Glykosylierung durchgeführt, die nun auch *O*-acetylierte und sulfatierte *N*-Glykane umfasst. Diese Analyse spiegelt wider, dass sulfatierte *N*-Glykane überwiegend an der IgA-Endstück-Region vorkommen, während *O*-acetylierte *N*-Glykane zehnmal seltener nachgewiesen werden als sulfatierte *N*-Glykane. Darüber hinaus wurden die MS-Daten der beiden IgA-Proben mit Hilfe von Oxonium-Marker-Ionen, die bei der eingehenden *N*-Glykoproteomanalyse von menschlichem Blutplasma beobachtet wurden, auch auf andere Modifikationen untersucht. Dies ermöglichte die ortsspezifische Identifizierung seltener *N*-Glykane mit Hexuronsäure (wahrscheinlich Glucuronsäure) als Capping, die an kontaminierende Proteine in beiden kommerziellen IgA-Proben gebunden sind.

Zusammenfassend lässt sich sagen, dass der entwickelte Arbeitsablauf geeignet ist, unsere Werkzeuge zur Beschreibung der *N*-Glykosylierung-Mikroheterogenität durch die Identifizierung seltener *N*-Glykane, die Aufklärung von *N*-Glykan-Strukturmerkmalen und den Nachweis von Glykoproteinen im mittleren bis niedrigen Konzentrationsbereich von Blutplasma-Proteinen zu erweitern. In Zukunft können die wichtigsten Merkmale des Datenanalyse-Prozesses auf Software-Tools für die Glykoproteomik übertragen werden, um eine genauere Interpretation von MS-basierten *N*-Glykoproteom-Messungen zu ermöglichen. Um die Möglichkeiten der Glykoproteomforschung zu erweitern, kann der vorgestellte Arbeitsablauf angepasst werden, um andere biologische Proben besser zu charakterisieren, z. B. Glykoproteine, die aus verschiedenen Organismen oder anderen Körperflüssigkeiten wie Urin, Milch, Speichel oder Schleimhautsekreten isoliert wurden.

Declaration

During the preparation of this thesis the author made moderate use of the online tools DeepL and Grammarly to check for additional adjustments related to grammar, clarity, and language. Nonetheless, the author reviewed and edited the content according the focus of the original ideas produced by the author. The author takes full responsibility for the content of the publication.

Magdeburg,

Francia Jaqueline Zúñiga-Bañuelos

Table of Contents

| | |
|---|------------------|
| Abstract | V |
| Kurzfassung..... | VIII |
| List of Abbreviations | XVII |
| | |
| <u>1</u> <u>Introduction</u> | <u>1</u> |
| | |
| <u>2</u> <u>Theoretical background</u> | <u>5</u> |
| | |
| 2.1. <u>Glycosylation.....</u> | <u>6</u> |
| 2.1.1. Protein <i>N</i> -glycosylation in humans | 6 |
| 2.1.2. The wide variability of <i>N</i> -glycan structures..... | 11 |
| 2.2. <u>Analysis of <i>N</i>-glycosylation of proteins in human blood plasma</u> | <u>15</u> |
| 2.2.1. Approaches for the study of glycoproteins..... | 16 |
| 2.2.2. Human blood plasma | 16 |
| 2.2.3. Standard methods applied for analyzing the <i>N</i> -glycosylation of blood plasma proteins | 17 |
| 2.2.4. In-depth analysis of the low-abundant <i>N</i> -glycoproteome | 19 |
| 2.2.4.1. <i>Sample preparation for low-abundant blood plasma proteins.....</i> | <i>19</i> |
| 2.2.4.2. <i>Liquid chromatography coupled to electrospray ionization with tandem mass spectrometry for</i> <i><i>N</i>-glycoproteomics</i> | <i>21</i> |
| 2.2.4.3. <i><i>N</i>-glycoproteomic data analysis.....</i> | <i>23</i> |
| 2.3. <u>Role of <i>N</i>-glycosylation in normal human physiology and pathophysiology</u> | <u>25</u> |
| 2.3.1. Normal physiology | 25 |
| 2.3.2. Human immunoglobulin A: a potential therapeutic glycoprotein..... | 26 |
| 2.3.3. Pathophysiology..... | 28 |
| 2.3.3.1. <i>Autoimmune diseases</i> | <i>30</i> |
| 2.3.3.2. <i>Congenital disorders of glycosylation (CDG).....</i> | <i>33</i> |
| 2.3.3.3. <i>Cancer</i> | <i>34</i> |
| 2.3.3.4. <i>Alzheimer's disease (AD)</i> | <i>34</i> |
| | |
| <u>3</u> <u>Materials and Methods.....</u> | <u>37</u> |

| | | |
|-------------|---|------------------|
| 3.1. | Reagents, human blood plasma, and IgA samples | 38 |
| 3.2. | Workflow for in-depth <i>N</i>-glycoproteomic analysis..... | 38 |
| 3.2.1. | Depletion of the top-14 high-abundant blood plasma proteins | 38 |
| 3.2.2. | GELFrEE® system fractionation | 39 |
| 3.2.2.1. | <i>Depleted blood plasma fractionation.....</i> | <i>39</i> |
| 3.2.2.2. | <i>IgA fractionation.....</i> | <i>39</i> |
| 3.2.3. | Proteolytic digestion..... | 40 |
| 3.2.3.1. | <i>Blood plasma-derived samples.....</i> | <i>40</i> |
| 3.2.3.2. | <i>IgA derived samples.....</i> | <i>41</i> |
| 3.2.4. | Glycopeptide enrichment..... | 41 |
| 3.2.4.1. | <i>Blood plasma-derived samples.....</i> | <i>41</i> |
| 3.2.4.2. | <i>IgA derived samples.....</i> | <i>41</i> |
| 3.2.5. | Glycoproteomic analysis of human blood plasma-derived samples..... | 42 |
| 3.2.5.1. | <i>nano RP-LC-ESI-OT-OT-MS/MS data-dependent acquisition.....</i> | <i>42</i> |
| 3.2.5.2. | <i>Data analysis</i> | <i>43</i> |
| 3.2.6. | Glycoproteomic analysis of human IgA samples..... | 46 |
| 3.2.6.1. | <i>nano RP-LC-ESI-OT-OT-MS/MS data-dependent acquisition.....</i> | <i>46</i> |
| 3.2.6.2. | <i>Data analysis on the fractionated IgA sample</i> | <i>47</i> |
| 3.2.6.3. | <i>Data analysis on two IgA commercial samples</i> | <i>48</i> |
| 4 | <u>In-depth analysis of the low-abundant <i>N</i>-glycoproteome of human blood plasma</u> | <u>53</u> |
| 4.1. | Introduction..... | 54 |
| 4.2. | Results..... | 56 |
| 4.2.1. | Exploring blood plasma (glyco-)proteins at low concentration range and <i>N</i> -glycan micro-heterogeneity | 57 |
| 4.2.2. | Development of a data analysis workflow for the identification of intact <i>N</i> -glycopeptides | 59 |
| 4.2.3. | Validated <i>N</i> -glycopeptide identifications in human blood plasma | 62 |
| 4.2.4. | Annotating structural features of bisecting, fucosylated, and LacNAc extended <i>N</i> -glycans using HCD.low spectra | 62 |
| 4.2.5. | Detection of sulfated and phosphorylated <i>N</i> -glycopeptides in common human blood plasma glycoproteins | 67 |
| 4.2.6. | Rare <i>N</i> -glycans in human blood plasma proteins..... | 70 |
| 4.3. | Discussion..... | 75 |
| 4.3.1. | Gain of the developed in-depth <i>N</i> -glycoproteomic analysis on human blood plasma | 75 |

| | | |
|----------|---|-------------------|
| 4.3.2. | Comparison of the performance of this workflow to other studies | 76 |
| 4.3.3. | Observations on the validation of <i>N</i> -glycopeptides with multiple fucoses | 77 |
| 4.3.4. | Relevance of <i>N</i> -glycan structure elucidation | 79 |
| 4.3.5. | Identification of glycopeptides bearing sulfated and phosphorylated <i>N</i> -glycans | 80 |
| 4.3.6. | Identification rare <i>N</i> -glycans in incorrectly-assigned gPSMs | 82 |
| 4.4. | Summary..... | 83 |
| 5 | <u>Site-specific identification of sulfated and <i>O</i>-acetylated <i>N</i>-glycans in human serum IgA.....</u> | <u>85</u> |
| 5.1. | Introduction | 86 |
| 5.2. | Results | 88 |
| 5.2.1. | Site-specific identification of FA2G2S2-SO ₄ <i>N</i> -glycan in the IgA heavy chain fraction | 89 |
| 5.2.1.1. | <i>Screening for MS² spectra containing HexNAc-sulfated oxonium marker ions</i> | <i>89</i> |
| 5.2.1.2. | <i>Glycoproteomic analysis focused on MS² spectra containing HexNAc-sulfated oxonium marker ions</i> | <i>90</i> |
| 5.2.2. | Expanded description of the <i>N</i> -glycosylation micro-heterogeneity of human serum IgA samples | 93 |
| 5.2.2.1. | <i>Site-specific identification of HexNAc-sulfated <i>N</i>-glycans.....</i> | <i>93</i> |
| 5.2.2.2. | <i>Site-specific identification of <i>O</i>-acetylated <i>N</i>-glycans.....</i> | <i>94</i> |
| 5.2.2.3. | <i>Peptide modifications detected in IgA <i>N</i>-glycopeptides.....</i> | <i>96</i> |
| 5.2.2.5. | <i>Identification of rare <i>N</i>-glycopeptides from “contaminant” proteins.....</i> | <i>96</i> |
| 5.2.2.6. | <i>Evaluation of the relative abundance of each IgA subclass</i> | <i>97</i> |
| 5.2.2.7. | <i>Expanded <i>N</i>-glycoproteomic analysis of IgA: integration of common and rare <i>N</i>-glycans</i> | <i>97</i> |
| 5.2.2.8. | <i>Comparison of two data analysis approaches for the identification of sulfated <i>N</i>-glycopeptides</i> | <i>101</i> |
| 5.3. | Discussion | 102 |
| 5.4. | Summary..... | 105 |
| 6 | <u>Conclusion and Outlook.....</u> | <u>107</u> |
| | <u>References</u> | <u>111</u> |
| | <u>List of Figures</u> | <u>127</u> |
| | <u>List of Supplementary Figures</u> | <u>128</u> |

| | |
|--|------------|
| A.1. In-depth analysis of the low abundant <i>N</i> -glycoproteome of human blood plasma | 128 |
| A.2. Site-specific identification of sulfated and <i>O</i> -acetylated <i>N</i> -glycans in human serum IgA | 130 |
| List of Tables | 131 |
| List of Supplementary Tables | 131 |
| A.1. In-depth analysis of the low abundant <i>N</i> -glycoproteome of human blood plasma | 131 |
| A.2. Site-specific identification of sulfated and <i>O</i> -acetylated <i>N</i> -glycans in human serum IgA | 132 |
| List of Materials and Chemicals | 133 |
| List of Instruments and Devices | 135 |
| Appendix | 136 |
| A.1. In-depth analysis of the low-abundant <i>N</i>-glycoproteome of human blood plasma | 137 |
| A.1. Tables | 137 |
| A.1. Figures | 137 |
| A.2. Site-specific identification of sulfated and <i>O</i>-acetylated <i>N</i>-glycans in human serum IgA | 159 |
| A.2. Tables | 159 |
| A.2. Figures | 159 |

List of Abbreviations

A

Ac-CoA
Acetyl-coenzyme A
Acetyl/Ac
Acetylation
ACPA
Anti-citrullinated protein antibodies
AD
Alzheimer's disease
AID
Autoimmune diseases
AUC
Area under the curve

C

C1q
Complement component
CDG
Congenital disorders of glycosylation
CHST
Carbohydrate sulfotransferase
CID
Collision-induced dissociation
CMP
Cytidine monophosphate
Cotton-HILIC-SPE
Cotton hydrophilic interaction liquid chromatography
solid-phase extraction

D

dhex
Deoxyhexose

E

ESI
Electrospray ionization
ETD
Electron transfer dissociation
ETHcD
Electron-transfer/higher-energy collision dissociation

F

FASP
Filter-aided sample preparation
FcγRIIIA
low affinity immunoglobulin gamma Fc receptor IIIa
FDR
False discovery rate
Fmoc
9-fluorenylmethyl chloroformate
Fuc
Fucose

G

GAGs
Glycosaminoglycans
Gal
Galactose
GalNAc
N-acetylgalactosamine
Glc
Glucose
GlcA
Glucuronic acid
GlcNAc

N-acetylglucosamine

GDP

Guanidine diphosphate

GPI

Glycosylphosphatidylinositol

gPSM

Glycopeptide spectrum match

H

HAP

Top 14 high-abundant blood plasma proteins

HCD

Higher-energy collision dissociation

HCD.low

HCD fragmentation with fixed NCE of 20

HCD.step

HCD fragmentation with stepped NCE of 20, 35, and 50

Hex

Hexose

HexA

Hexuronic acid

HexNAc

N-Acetylhexosamine

HILIC

Hydrophilic interaction liquid chromatography

HILIC-UHPLC-FLD

Hydrophilic-interaction ultra-high performance liquid chromatography with fluorescence detection

HNK-1

Human natural killer-1

HPLC

High-performance liquid chromatography

I

IgA

Immunoglobulin A

IgG

Immunoglobulin G

IMforFUTURE

Innovative training in methods for future data

L

LacNAc

N-Acetylactosamine

LC-ESI-MS/MS

Liquid chromatography coupled to electrospray ionization with tandem mass spectrometry

LC-MS/MS

Liquid chromatography with tandem mass spectrometry analysis (in general terms)

M

m/z

Mass to charge ratio

mAb

Monoclonal antibody

MALDI-TOF-MS

Matrix-assisted laser desorption/ionization time-of flight mass spectrometry

Man

Mannose

MIRAGE

Minimum Information Required for A Glycomics Experiment

MPI

Max Planck Institute for Dynamics of Complex Technical Systems

MS

Mass spectrometry

MS/MS

Two-step mass spectrometry technique

N

nanoLC

Nanoscale liquid chromatography

nanoRP-LC-ESI-OT-OT-MS/MS

Nano-reversed-phase liquid chromatography electrospray-ionization orbitrap tandem mass

spectrometric measurement for precursor and
fragment ions

NCE

Normalized collisional energies

NeuAc

5-N-Acetyl-neuraminic acid

O

OST

Oligosaccharyltransferase

P

PAPS

3'-Phosphoadenosine-5'-phosphosulfate

PGC

Porous graphitized carbon

Phospho/P

Phosphorylation

Protein Abundance Database

PaxDB

R

RP

Reversed-phase stationary phase

S

SDS-PAGE

Sodium dodecyl sulfate-polyacrylamide gel
electrophoresis

Sulfo/S

Sulfation

U

UDP

Uridine diphosphate

X

xCGE-LIF

Multiplexed capillary gel electrophoresis with laser-
induced fluorescence detection

XIC

Extracted ion chromatograms

1 Introduction

This thesis emerged from the program “Innovative Training in Methods for Future Data (IMforFUTURE),” a Marie Skłodowska-Curie Innovative Training Network. This program aimed to develop methods, produce datasets and apply integrative data analyses that contribute to comprehending biological processes influencing diseases and aging. IMforFUTURE funded an interdisciplinary consortium of experts in data analysis and multiple omics such as glycomics, glycoproteomics, epigenomics, and genomics. This consortium included glyXera GmbH, an industry partner specialized in glycomics. The collaboration was extended to the Bioprocess engineering department of the Max Planck Institute (MPI) for Dynamics of Complex Technical Systems, where the Bio/Process-Analytics team specializes in developing methods for analyzing protein glycosylation. All this motivated the establishment of a workflow that serves the purposes of an in-depth *N*-glycoproteomic analysis of human blood plasma and the production of new datasets.

Conducting an in-depth *N*-glycoproteomic analysis on low-abundant human blood plasma proteins is advantageous for several reasons. First, glycans (carbohydrate chains) are cellular building blocks as essential as proteins, lipids, and nucleic acids. The importance of glycans as a macromolecular component of cells was recognized since it is impossible to understand the origin of many diseases (e.g. inflammatory bowel disease, IgA nephropathy, and rheumatoid arthritis) without considering glycans in a site-specific manner [1, 2]. Alterations of *N*-glycosylation in human proteins have been observed not only on various diseases [3-8], but also through physiological conditions, such as pregnancy and aging, by analyzing blood plasma or serum samples [7, 9-11]. Second, the analysis of *N*-glycans offers extensive information not only exclusive from *N*-glycosylation but also general to glycoconjugates biosynthesis, and their site-specific identification is less challenging than for the other glycan types. Third, proteins have multiple functions and, within a complex organism, several are present in a tissue-specific manner. Tissue specific proteins can be identified in the low-abundance concentration range in human blood plasma. In addition, although proteins are derived from a nucleic acid-based template, the proper folding and function of several proteins depend on the linkage of one or more *N*-glycans. It is known that *N*-glycosylation is present in around 45% of the proteins from the eukaryotic domain and it is crucial for several biological functions (e.g., cell interactions and development) [12]. Thus, beneficial for diagnostics and therapeutics, the low-abundant glycoproteome could reveal information associated with pathological processes in cell tissues or an impaired metabolism at an early stage of the disease [5, 13].

One challenging feature of glycosylation is that this complex post-translational process is non-template-based. The enormous diversity of *N*-glycan structures attached to a protein was quoted for the first time as glycan heterogeneity in 1961 [14, 15]. This implies that if one protein has four *N*-glycosylation sites and ten different *N*-glycans may be attached to each site, then 10,000 different protein glycoforms may be observed. This number of combinations is narrowed down by the intracellular microenvironment, protein structure, enzyme gene expression, and other factors [16-23]. An introduction to protein glycosylation, *N*-glycan biosynthesis, and the structural diversity of *N*-glycans in humans is presented in chapter 2, section 2.1. Another challenge arises when studying low-abundant glycoproteins from human blood plasma. Blood is a widespread circulatory fluid that networks organs and tissues, a high diversity of proteins is present in the blood plasma [24]. However, the blood plasma proteins are spread in a concentration range of at least ten orders of magnitude [24-26]. Previous studies have shown that tissue-leakage proteins exist in the low-concentration range of blood plasma proteins [27]. In addition, probably also at low levels within the *N*-glycosylation profile of proteins, some rare *N*-glycan structures are not well described in a site-specific manner [28, 29]. All in all, the study of the low-abundant glycoproteome from blood plasma is not only hindered by proteins of higher abundance but also by the number of glycoforms per protein [30].

To reduce suppression of precursor ions signals from low-abundant glycopeptides, the steps of a sample preparation workflow must include protein or peptide fractionation and a bottom-up mass spectrometry (MS) approach using a high-performance mass spectrometric acquisition [31]. Then, to conduct an *N*-glycoproteomic data analysis, a software-assisted search is necessary to automate the site-specific identification of *N*-glycans, followed by a data validation process. In chapter 2, section 2.2. a summary of the current analytical methods for studying *N*-glycosylation from human blood plasma proteins shows their advantages and disadvantages. This will strengthen the rationalization behind developing innovative in-depth glycoproteomic technologies, such as the workflow established in this thesis. Finally, the section 2.3. will explain the importance of studying alterations of *N*-glycosylation in normal physiological states and pathologies that show increasing trends of global incidence without adequate treatments. An overview of the diagnosis, disease development, and treatment of congenital disorders of glycosylation (CDG) [32-34], and other prevalent diseases like autoimmune diseases (AID) [35-41], cancer [5, 13, 42], and Alzheimer's disease (AD) [43], are given to contextualize the potential contribution of further research in the field of glycoproteomics. Thus, chapter 2 provides concepts and arguments for the contextualization of the subject matter of this thesis.

The first objective of this thesis was to develop a workflow for the in-depth *N*-glycoproteomic analysis of human blood plasma proteins, which must achieve the characterization of intact *N*-glycopeptides and the

site-specific identification of *N*-glycans within the low-abundant *N*-glycoproteome of human blood plasma. For this purpose, the *N*-glycoproteomic workflow had to overcome two drawbacks: detecting the low-abundant blood plasma proteome and reliably describing the structure of *N*-glycans on a protein *N*-glycosylation site manner using MS. Therefore, after screening various fractionation approaches, the first drawback was counterbalanced by establishing a sample preparation workflow that features the depletion of fourteen high-abundant blood plasma proteins followed by the fractionation of the depleted fraction via a GELFrEE® system. The sample preparation workflow had to incorporate proteolytic digestion and an improved glycopeptide enrichment to adapt the samples for MS analysis. Applied for the fragmentation of intact glycopeptides, the two MS-based measurements implemented are differentiated by the normalized collisional energies (NCE) used: stepped (NCE 20, 35, 50%) and fixed low (NCE 20%). The second drawback was overcome with a data analysis workflow that includes the application of a decision tree for data validation and integrating both MS measurements to augment the description of intact *N*-glycopeptides. Thus, the workflow for the in-depth *N*-glycoproteomic analysis of human blood plasma proteins include a sample preparation and a data analysis workflow. The methodology describing both workflows is presented in chapter 3.

Significant gains resulted by applying the established *N*-glycoproteomic workflow to a blood plasma standard sample. First, more glycoproteins about the middle- down to low-abundant blood plasma concentration range were detected. Second, rare *N*-glycans (e.g., glucuronidated, sulfated, phosphorylated, and other rare *N*-glycan building blocks), corresponding to glycoproteins within the middle-abundant concentration range, were identified. Their *N*-glycan structures were partially described. Third, rare *N*-glycan structures were elucidated via the integration of data from two previously established fragmentation methods. The integrated data revealed the presence of an *N*-glycan structure featuring a rare glycosidic linkage between a sialic acid and an *N*-acetylhexosamine (HexNAc), which is assumed to be nonexistent in human *N*-glycans. Fourth, the data analysis workflow identified potential improvements for glycoproteomic software design and data validation strategies. These results are presented and discussed in chapter 4.

The second objective of this thesis was to use the developed *N*-glycoproteomic workflow for the site-specific identification of a protein *N*-glycosylation feature that has been difficult to describe by current glycoproteomic approaches. This is the case for *N*-glycans bearing sulfated HexNAc in human serum immunoglobulin A (IgA), a human serum glycoprotein of therapeutic interest. Sulfated *N*-glycans from mammalian IgA were detected for the first time by Boisgard *et al.*, and only after 20 years this was confirmed in *N*-glycans released from human serum IgA by glycomic analyses conducted in our group [28, 44, 45]. However, current *N*-glycoproteomics experiments have not yet detected glycopeptides bearing

sulfated *N*-glycans in IgA. Hence, sulfated *N*-glycans have not been assigned to specific *N*-glycosylation site(s). Motivated by this challenge, the newly developed in-depth *N*-glycoproteomic workflow was applied to identify intact sulfated *N*-glycopeptides derived from human IgA. Then, the data analysis workflow was optimized. This optimization enabled the site-specific identification of a sulfated *N*-glycan previously observed in our group and revealed other *N*-glycan compositions on IgA holding sulfated HexNAc. After this first identification of sulfated *N*-glycopeptides, two commercial IgA samples were analyzed using a more straightforward sample preparation but applying the optimized *N*-glycoproteomic data analysis workflow. The application of the new data analysis workflow was extended to the search for other rare *N*-glycans. It was identified that *N*-glycosylation sites IgA bear *O*-acetylated sialic acid at several sites and that contaminant glycoproteins bear *N*-glycans with terminal glucuronic acid. In addition, the low-abundant *N*-glycopeptides bearing sulfated HexNAc were reproducibly identified primarily at the tailpiece *N*-glycosylation site of IgA. This outcome demonstrated that our newly developed *N*-glycoproteomic workflow is adaptable to other samples and enables the detection of these difficult-to-identify *N*-glycopeptides. These results are presented in chapter 5.

In summary, the application of the developed workflow for in-depth *N*-glycoproteomic analysis enables explorative studies, supports the comprehensive identification of low-abundant *N*-glycopeptides, and allows the site-specific *N*-glycan identification using the oxonium ion-guided interpretation of spectra without prior knowledge of the peptide sequence. It is essential to recognize the technical challenges of this workflow, which are the costly resources required, the long time needed for implementation, including data interpretation, and the bioinformatics resources necessary to handle the enormous amount of MS data to be analyzed. Nevertheless, the datasets produced by our *N*-glycoproteomic workflow include the observation of data derived from the low-abundant glycoproteome. Thus, our data is rich in glycopeptides rarely evaluated during screening biomarkers or therapeutic options. The final goal of the IMforFUTURE project is to apply this workflow to the production of datasets from a clinical cohort of interest. At the time of writing this thesis, this goal remains a prospect. On a prospective basis, generating such valuable datasets could contribute to understanding pathophysiological processes related to autoimmune disease and improving diagnostic and therapeutic tools.

2 Theoretical background

Glycosylation is a post-translational modification in which a carbohydrate side chain is enzymatically linked to an aglycone – e.g., a nascent protein or lipid [46]. As Figure 2.1 shows, this modification results in different types of molecules like glycoproteins, proteoglycans and glycosphingolipids.

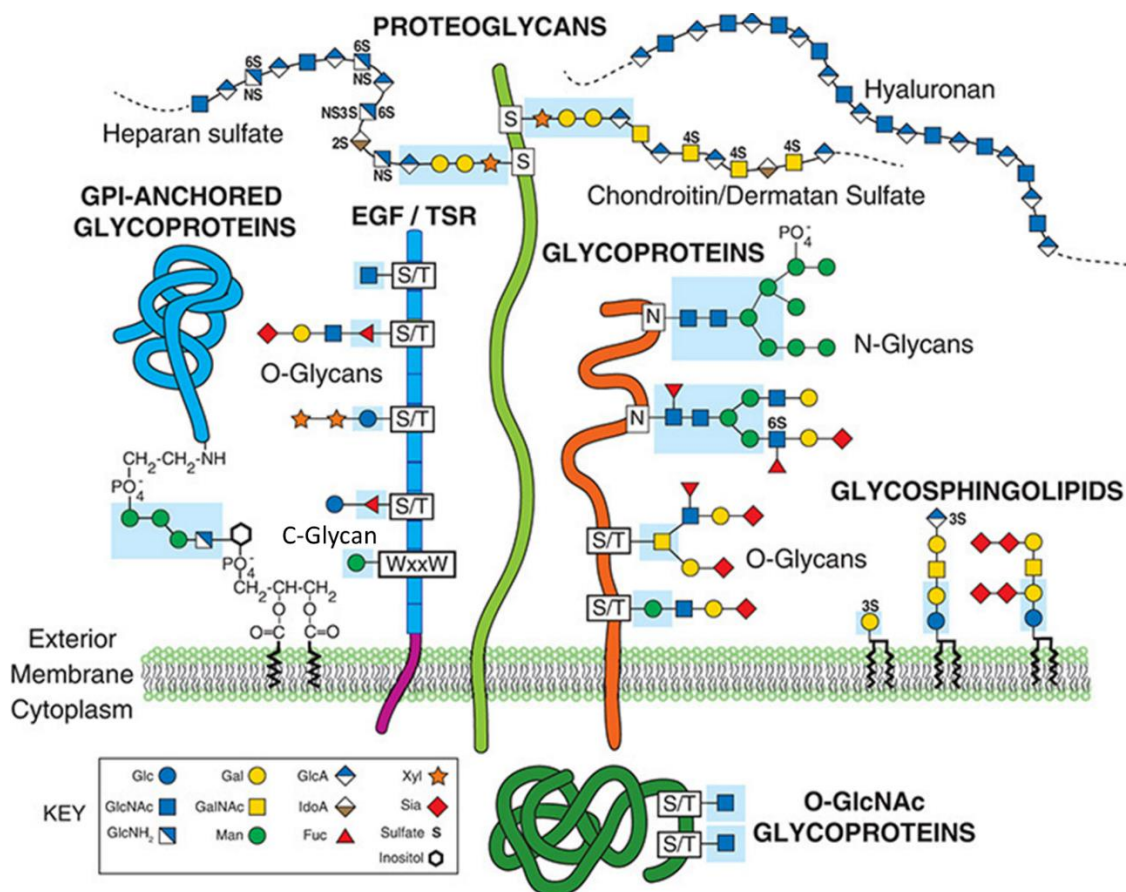


Figure 2.1. Diversity of glycoconjugates in animal cells. Modified from Varki and Kornfeld [47]. The oligosaccharides linked to lipids result in two subclasses, glycosphingolipids and glycosylphosphatidylinositol (GPI). Glycosphingolipids have two components a ceramide lipid and a glycan that can be relatively diverse [48]. GPI is more complex and has three components: 1) a lipid linked with a phosphate group to an inositol sugar (phosphatidylinositol moiety), 2) a conserved carbohydrate of four monosaccharides, and 3) an ethanolamine phosphate moiety [49]. The last moiety serves as a bridge for attaching GPI to the C-terminal domain of a protein, which is anchored to the cell membrane. Hyaluronan is a secreted long carbohydrate chain with approximately 80 non-sulfated monosaccharides that belongs to the glycosaminoglycans (GAGs) subclass [50]. In general, GAGs are composed of repeated disaccharide units (an amino sugar and a uronic acid), which can be modified with sulfation (e.g., heparan sulfate). When GAGs are covalently attached to a protein, the resulting glycoconjugate is named proteoglycan. GAGs and proteoglycans are a major constituent of the extracellular matrix [50]. Finally, glycoproteins are proteins carrying O-, C-, or N-linked glycans attached to serine/threonine, tryptophan, or asparagine, respectively.

2.1. Glycosylation

Carbohydrates associated with a protein, now called glycans, were chemically detected for the first time on ovalbumin in 1893, but this discovery was confirmed only in 1938 by Neuberger [13, 14]. Glycans can be attached to proteins with three linkages. Depending on the amino acid atom with which the first sugar is covalently attached it might result in *N*-, *C*-, or *O*-linked glycans [51-54]. *N*-linked glycans are covalently bonded through a β -linkage to the nitrogen from an asparagine side chain. *N*-glycosylation occurs specifically at the protein sequence: Asn-X-Ser/Thr (being X any amino acid except Pro) [55]. *N*-glycosylation is the aim of this thesis and the following sections in this chapter will describe it further. In regard to *C*-linked glycans, only *C*-linked mannose is known [56]. *C*-mannosylation occurs when a mannose is covalently attached in an α -linkage to the indole carbon of a tryptophan belonging to the amino acid sequence: Trp-X-X-Trp/Cys (where X is any amino acid) [57]. *O*-glycans are covalently bonded to the oxygen of a serine/threonine's hydroxyl group of a nascent protein [56]. *O*-glycosylation, in contrast to the other glycosylation types, occurs in an unpredictable fashion, resulting in multiple and consecutive *O*-glycosylation sites [58]. In humans, the subclasses of *O*-glycans include *O*-fucose, *O*-glucose, *O*-mannose, *O*-linked *N*-acetyl glucosamine (*O*-GlcNAc), and *O*-linked *N*-acetyl galactosamine (*O*-GalNAc)—named after the first sugar *O*-linked to the protein [56]. The most common *O*-glycosylation is the α -linked *O*-GalNAc, which frequently generates mucin type glycans. Due to the unpredictable position of *O*-glycans and challenging characterization of *C*-mannosylated proteins, the study of *O*- and *C*-glycosylation is preferred yet on targeted proteins and not in complex samples.









Widespread in all life domains, the biosynthesis of glycoconjugates in eukaryotic cells diverges significantly from bacteria and archaea [58, 59]. While in bacteria, glycosylation is necessary for pathogenesis but not development, in eukaryotic cells, it is crucial for cell development, regulatory processes, and protein folding, etc. [60]. However, the complex biosynthesis pathway of these oligosaccharides in eukaryotic organisms still presents a tremendous diversity among plants, fungi, and animals and even diverges between animal cells [46, 58, 59]. Thus, defining the organism to study and its associated glycosylation features (e.g., *N*-glycan building blocks) is an initial step when analyzing its *N*-glycans. As the scope of this thesis is the analysis of *N*-glycosylation in human blood plasma proteins, specific details of protein *N*-glycosylation in humans are introduced in the following section.

2.1.1. Protein *N*-glycosylation in humans

The eight most common building blocks of human *N*-glycans plus three molecules that add post-glycosylation modifications are displayed in Table 2.1, together with the corresponding nucleotide sugar

donors used as a substrate of the glycan biosynthesis [61, 62]. These *N*-glycan building blocks are condensed into five groups by their chemical formula: *N*-acetylhexosamines, hexoses, deoxyhexoses, uronic acids, and nonulosonic acids. Uronic acids are acidic sugars holding a carboxylic acid at 6-carbon, which are usually derived from hexoses and known as hexuronic acids (HexA) [47]. One typical example of an hexuronic acid is glucuronic acid (GlcA), an elementary building block of GAGs but also reported in *N*-glycans of the nervous system [50, 63]. Nonulosonic acids are a class of complex 9-carbon backbone monosaccharides that carry diverse chemical groups on the side chains (i.e., one carboxylic acid and one or multiple acetylated amino groups) and even modifications such as sulfation or methylation [64]. The most common nonulosonic acid in humans is 5-*N*-acetyl-neuraminic acid with a characteristic *N*-acetyl group at the 5-carbon (Neu5Ac) – referred as NeuAc in this thesis. Frequently, this terminology appears in the identification of *N*-glycans by MS, as only mass-to-charge ratios (m/z) and not molecular configurations are detected via MS.

Table 2.1. Typical *N*-glycan building blocks in humans.

| <i>N</i> -glycan building block | Abbreviation | Nucleotide sugar donor | Elemental composition (dehydrated) | Theoretical Mass (Da) | Symbol |
|--|--------------|------------------------|--|-----------------------|---|
| <i>N</i> -acetylglucosamine ¹ | GlcNAc | UDP-GlcNAc | C ₈ H ₁₃ O ₅ N | 203.0794 |  |
| <i>N</i> -acetylgalactosamine ¹ | GalNAc | UDP-GalNAc | C ₈ H ₁₃ O ₅ N | 203.0794 |  |
| Galactose ² | Gal | UDP-Gal | C ₆ H ₁₀ O ₅ | 162.0528 |  |
| Mannose ² | Man | GDP-Man | C ₆ H ₁₀ O ₅ | 162.0528 |  |
| Glucose ² | Glc | UDP-Glc | C ₆ H ₁₀ O ₅ | 162.0528 |  |
| Neuraminic acid ³ | NeuAc | CMP-NeuAc | C ₁₁ H ₁₇ O ₈ N | 291.0954 |  |
| Fucose ⁴ | Fuc | GDP-Fuc | C ₆ H ₁₀ O ₄ | 146.0579 |  |
| Glucuronic acid ⁵ | GlcA | UDP-GlcA | C ₆ H ₈ O ₆ | 176.0321 |  |
| Phosphorylation | Phospho | GlcNAc-1-P | H ₁₀ O ₃ P | 79.9663 | P |
| Sulfation | Sulfo | PAPS | O ₃ S | 79.9568 | S |
| Acetylation | Acetyl | Ac-CoA | C ₂ H ₂ O ₁ | 42.0106 | Ac |

PAPS: 3'-Phosphoadenosine-5'-phosphosulfate, Ac-CoA: acetyl-coenzyme A, UDP: Uridine diphosphate, CMP: cytidine monophosphate, GDP: guanine diphosphate. The numbers in superscripts identify the five groups of *N*-glycan building blocks. 1: *N*-acetylhexosamine (HexNAc), 2: Hexose (Hex), 3: Nonulosonic acid, 4: Deoxyhexose (dhex), and 5: Uronic acid.

The protein *N*-glycosidic bond occurs between the amide nitrogen of the L-asparagine side chain and the reducing end of an *N*-acetyl- β -D-glucosamine [65]. Structurally stable, a glycosidic bond is a linkage between the reducing end of an (oligo) saccharide and the hydroxyl group of a C-atom in another sugar (or amide nitrogen). The reducing end of a sugar moiety consists of an anomeric carbon (typically carbon 1) holding a hydroxyl group in either α or β anomeric configurations (down axial orientation or up equatorial direction, respectively) [65]. As shown in Figure 2.2, the nomenclature of the glycosidic linkages is described by the anomeric conformation (α - and β -glycosidic linkage) and the numbers of both the anomeric carbon and the C-atom of the linked sugar (e.g., β 1-2, β 1-4)[65]. The composition and glycosidic bonds between the initial five saccharides of the *N*-glycan are always conserved. This *N*-glycan-core

structure comprises two GlcNAc followed by one mannose that links two branching mannoses: one on the C3 and another on the C6 of the first mannose [53]. After the core-structure, the glycosidic bonds in the branches (so called antennae) of the *N*-glycan present more diversity. Depending on the monosaccharide composition of the antennae, the *N*-glycan is classified into three main groups: complex-, hybrid-, and oligomannose-type *N*-glycan. Their biosynthesis is briefly explained below.

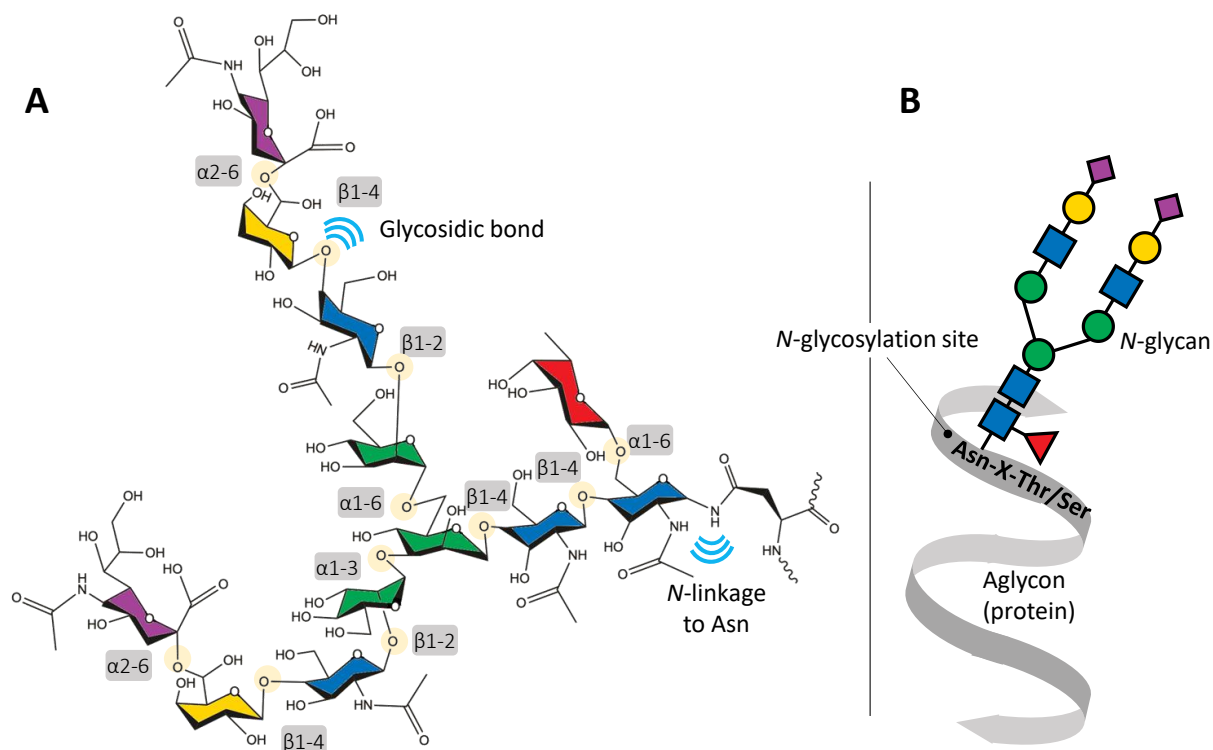


Figure 2.2. Protein *N*-glycosylation. A) Detailed description of the chemical structure of a complex-type *N*-glycan. The glycosidic bonds between each *N*-glycan building block are described in gray boxes near the linkages. Blue arcs illustrate the two linkage-types: glycosidic bonds (linkage between monosaccharides highlighted in light beige) and N-linkage to the nitrogen of an asparagine side chain. Wiggly black lines indicate a continuation of the protein amino acid chain. B) Primary elements of protein *N*-glycosylation. *N*-glycosylation site: Asn-X-Thr/Ser amino acid sequence. Adapted from van Kooyk and Rabinovich [66].

As displayed in Figure 2.3, *N*-glycosylation starts in the endoplasmic reticulum (ER), where a precursor *N*-glycan chain, composed of two GlcNAc, nine mannoses, and three glucoses, is transferred from a lipid (dolichol phosphate) to a nascent protein [53]. Further *N*-glycan processing in the ER enables a protein-folding quality control dependent on glucose removal during protein translation. The glycoprotein folding process ends with an attached *N*-glycan of two GlcNAc and eight mannoses. Then, the glycoprotein is exported to the cytoplasm (as oligomannose-type *N*-glycan) or the Golgi apparatus. The mannosylated *N*-glycan is trimmed and modified by glycosidases and glycosyltransferases anchored along specific Golgi cisternae (*cis*-, *medial*-, and *trans*-Golgi). For example, the *N*-acetylglucosaminyltransferase 1 (MGAT1) anchored in the medial-Golgi transfers to the *N*-glycan one branching GlcNAc. This allows the synthesis of

hybrid-type *N*-glycans and the activity of α -mannosidase II enzymes, which enable the formation of complex-type *N*-glycans. Other examples are galactosyltransferases and sialyltransferases, located in the *trans*-Golgi, which diversify further in hybrid- and complex-type *N*-glycans.

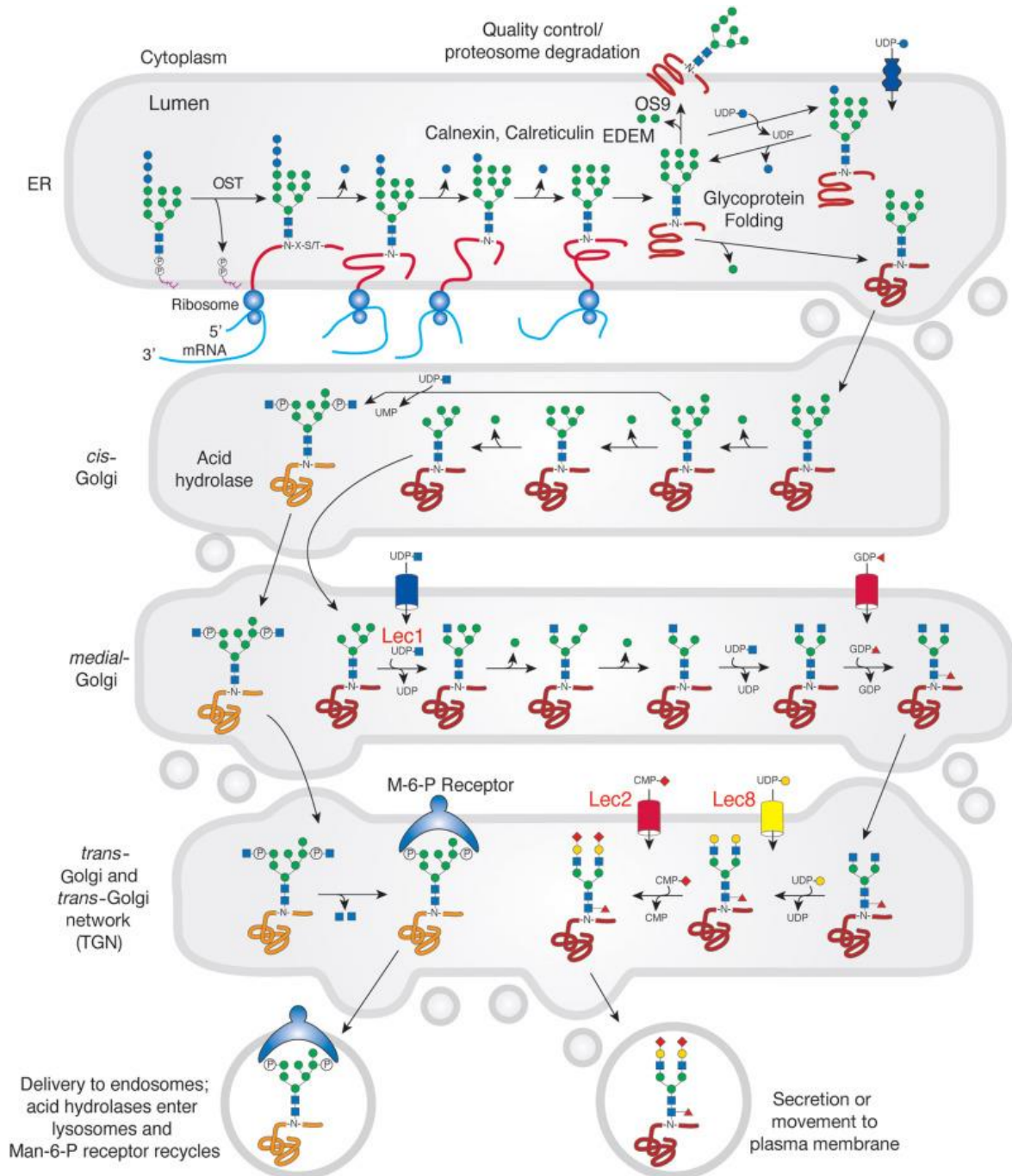


Figure 2.3. *N*-glycan processing pathway. Taken from Stanley et al. [53]. The endoplasmic reticulum (ER) is presented at the top, where the *N*-glycan Glc3Man9GlcNAc2 attached to the lipid dolichol pyrophosphate is matured. The oligosaccharyltransferase (OST) enzyme transfers the matured *N*-glycan to the *N*-glycosylation site of the translated protein. Chaperons, like calnexin and calreticulin, or the UDP-glucose: glycoprotein glucosyltransferase (UGGT), interact with the protein while glucosidases and one

mannosidase remove three glucose and one terminal mannose as a signal of correct protein folding. Misfolded proteins are processed by ER degradation-enhancing α -mannosidase I-like proteins (EDEMs) and then directed to degradation in the cytoplasm by the lectin OS9. The Golgi apparatus is represented below the ER with three cisternae: cis-Golgi, medial-Golgi, and trans-Golgi. After the correct protein folding, several proteins carrying N-glycan(s) with eight or nine mannoses can be transported to the cis-Golgi, where additional mannoses can be removed. In the medial-Golgi, the MGAT1 transfers the first GlcNAc at the C-6 Mannose of the Man5GlcNAc2, producing hybrid N-glycans. α -Mannosidase II removes two terminal mannoses at the C-3 Mannose, which is the substrate for the N-acetylglucosaminyltransferase MGAT2 to produce the N-glycan GlcNAc2Man3GlcNAc2. The biantennary N-glycan is further modeled by adding fucose, galactose, and neuraminic acid to produce complex-type N-glycans. Bisecting GlcNAc, additional antennae, different capping sugars, antennae or core fucosylation, sulfation, and polyLacNAc extensions can occur on the complex type N-glycans. Many of these processes can be blocked by the Lec transcription factors. In the case of lysosomal hydrolase proteins, the N-glycans are decorated by GlcNAc-1-P at C-6 of Man residues in the cis-Golgi. The GlcNAc is cleaved in the trans-Golgi by a phosphodiester N-acetylglucosaminidase (NAGPA), and the phosphorylated mannoses are recognized by Man-6-P receptor, which transports this protein to lysosomal vesicles. The arrows indicate the regular processes, movements, and enzymatic reactions. The activated sugars UDP-Glucose, UDP-GlcNAc, UDP-Galactose, GDP-Fucose, and CMP-Neuraminic acid are represented with respective letters, and the sugars are symbolized with the respective figures and colors.

In summary, oligomannose-type N-glycans, secreted by the ER or the cis-Golgi, bear branches extended by mannose units. As shown in Figure 2.3, the complex- and hybrid-type N-glycan structures must transit at least until the medial-Golgi. Complex-type N-glycans typically end up with the maximum maturation by reaching the trans-Golgi, where the N-glycan antennae are elongated with capping, galactose, and sialic acid or more antennae (generated through branching GlcNAc). Hybrid-type N-glycan structure combines an oligomannose-type branch on the C6-linked mannose and a complex-type branch on the C3-linked mannose. Complex- and hybrid-type N-glycans may also present fucose linked primarily on the core GlcNAc and less frequently on the antenna galactose or GlcNAc. A high diversity in glycosidic bonds is characteristic of complex and hybrid N-glycans. However, the possible combinations for an N-glycan structure can be narrowed down by considering the human N-glycan biosynthetic pathways. For example, proper from humans, the N-glycan in the core-structure might have a glycosidic linkage between one fucose and the first GlcNAc in an α 1-6 linkage but not α 1-3, the last is proper of insect cells [59].

Other processes that can modify the hydroxyl groups from N-glycans are phosphorylation, acetylation, and sulfation [46]. Mannose phosphorylation occurs in the cis-Golgi by the transference of GlcNAc-1-P (from UDP-GlcNAc) to the carbon six (C6) of a Man by a phosphotransferase [67]. Then the GlcNAc is removed by a phosphodiester glycosidase in the trans-Golgi, producing N-glycans bearing Man6P [53]. O-acetylation of sialic acids occurs most commonly at C7 and C9 [64]. The sugar nucleotide donor CMP-Neu5Ac receives an acetyl group from acetyl-CoA by an O-acetyltransferase (e.g., CASD1) in the trans-Golgi [64]. Regarding sulfated glycans, sulfate groups modify GlcNAc, Gal, GalNAc, and GlcA, primarily [68]. These sulfated monosaccharides are typical of GAGs and atypical of N-glycans [68]. Sulfation of the epitope sialyl Lewis x, present in N-glycans, is carried in the trans-Golgi by the carbohydrate sulfotransferase 1 (CHST1) at the C6 of Gal or by CHST2 at the C6 of GlcNAc [53, 69]. Sulfation at the C4 of terminal GalNAc is found in N-glycans linked to glycoprotein hormones. Typical of brain O- and N-glycans, sulfation at the C3 of GlcA from the

human neutrophil killer carbohydrate (HNK-1) is catalyzed by the CHST10 [69]. The transference of GlcA to Gal is catalyzed by the glucuronyltransferases GlcAT-P (B3GAT1) and GlcAT-S (B3GAT2)—based on mice experiments, GlcAT-P and GlcAT-S activity is selective from brain and kidney, respectively [70-72].

N-glycan maturation is multifactorial and, so far, unpredictable. The synthesis of this *N*-linked oligosaccharide depends on several variables. The influence power of each variable differs for each protein and *N*-glycosylation site. The first variable is the protein structure and the position of each site on the protein [73]. Second, *N*-glycan expression is influenced by the homeostasis in the Golgi. The activity of enzymes and trafficking molecules in the Golgi is affected by the pH, oxidative potential, and Cl⁻, Ca²⁺, Mn²⁺, and K⁺ ions balance [17]. For example, hypoxia inhibits the hetero-dimerization of sialyltransferases with galactosyltransferases [18]. Third, *N*-glycan structure is directly affected by the expression levels of glycosyltransferases and glycosidases, which are regulated by genetics, epigenetics, and transcriptomics (including regulatory microRNAs) [19-21]. Fourth, enzyme-substrate availability modifies the type of *N*-glycan structures synthesized. This was specially studied in the production of recombinant glycoproteins where deficiencies in the medium formulation impair the availability of nucleotide sugar donors [74]. Another example is shown by the severe effect of the congenital disorders of glycosylation of which the synthesis of an activated sugar (e.g., GDP-fucose) is scarce [3]. Fifth, *N*-glycosylation will be affected by the impaired function and expression of other essential Golgi component proteins such as cargo proteins, ion pumps, ion channels, chaperons, and trafficking molecules [17, 22, 23]. This explains how the complexity of an *N*-glycosylation profile captures the microenvironment in a cell, tissue, organism, or cohort at a defined time point. However, although *N*-glycan synthesis is unpredictable, each *N*-glycosylation site displays a consistent distribution of *N*-glycan structures if the environment remains constant (in a steady state). Thus, when multiple environments are compared, e.g., suspension vs adherent cells, alterations of *N*-glycosylation display not only differences among any of the variables mentioned above but also opportunities for modulating protein *N*-glycosylation [75]. Especial microenvironments are also reflected during health and disease, e.g., differences in the *N*-glycosylation of anti-citrullinated protein antibodies (ACPA) and not simultaneously detected in the total serum IgG at the onset of rheumatoid arthritis [76]. This thesis aims to enable and expand the extraction of site-specific and structural *N*-glycosylation data from human blood plasma proteins with a special interest in including the low-abundant protein glycoforms.

2.1.2. The wide variability of *N*-glycan structures

Expression of a protein-coding gene will always result in the same polypeptide sequence. If the polypeptide sequence is *N*-glycosylated, the *N*-glycan structure gives the protein a different structural form called “glycoform.” While one *N*-glycosylation site can receive different *N*-glycans, one protein can also have

multiple glycosylation sites [77]. This results in a complex mix of protein glycoforms that can be described at two levels of diversity: micro-heterogeneity (variability of *N*-glycans per site) and macro-heterogeneity (variability of protein sites occupied with *N*-glycans). Estimations indicate that *N*-glycosylation occurs at approximately 66% of the predicted sites with an occupational preference for *N*-glycosylation sites surrounded by hydrophobic amino acids [77, 78]. Studies related to protein macro-heterogeneity show that *N*-glycosylation tends to occur in the proximity of turns and bends of the protein secondary structure, leading to a decrease in the protein folding freedom and, thus, controlling its structural stability [78, 79].

Multiple *N*-glycosylation sites on a protein might play different roles. For example, human diamine oxidase, a key enzyme for histamine catabolism, has four *N*-glycosylation sites. Essential for dimer formation, the *N*-glycosylation sites N538 and N745 of diamine oxidase bear complex-type *N*-glycans. In contrast, bearing oligomannose-type *N*-glycans, the sites N168 and N110 are critical for protein folding and expression [16]. This demonstrates the crucial site-specific function of the micro-heterogeneity of *N*-glycosylation, encouraging the analysis of protein *N*-glycosylation. Nonetheless, to describe the micro-heterogeneity of *N*-glycosylation, it is first essential to understand the structural complexity levels of an *N*-glycan, depicted in Figure 2.4.

Considering the symbology presented in Figure 2.4.A, the five *N*-glycan structural complexity levels that increase conformational diversity (Figure 2.4.B.) are described as follows:

- Composition: the number and variety of monosaccharides linked
- Post-glycosylational modifications: additional chemical groups linked to the monosaccharides
- Branching: the not linear but tridimensional structure given by the branching nodes
- Connectivity: the specific order of the monosaccharides connected by glycosidic bonds
- Anomeric configuration: the orientation of the OH groups that formed the glycosidic linkages

Shown in Figure 2.4.C, the levels of structural complexity significantly expand the diversity of the main *N*-glycan groups [46]. For example, phosphorylation and the number of mannoses can diversify the oligomannose-type *N*-glycans. In hybrid-type *N*-glycans at least one of the antennae can vary by the number and linkages of other monosaccharides like fucose, sialic acid, and the presence of modifications such as sulfation. The structural diversity increases for human complex-type *N*-glycans because, in addition to the previous complexities, multiple antennae, various fucose positions, sialic acid *O*-acetylation, bisecting GlcNAc, elongations of *N*-acetylactosamine units (LacNAc), and other special modifications, are possible at any of both branching mannoses.

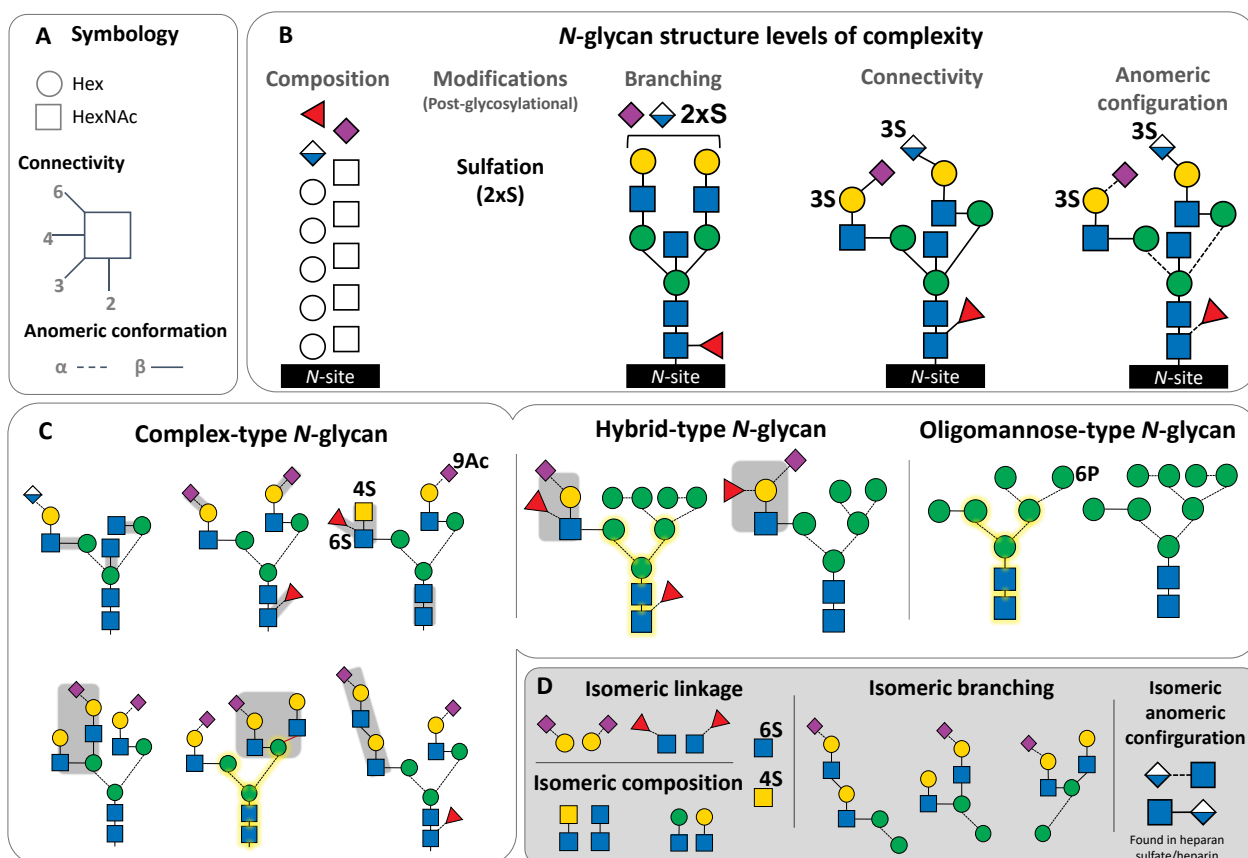


Figure 2.4. Structural complexity of *N*-glycans. *N*-glycan common features are named below or above each structure. Gray shade indicates isomeric potential fragment ions. Bisecting GlcNAc is highlighted in yellow. *N*-glycan core-structure highlighted in yellow. Modified from Gabarics et al. [80].

Hard for deciphering via MS are a share of *N*-glycan structural features that belong to isomeric structural fragments—molecules with the same chemical formula but different spatial arrangement [80]. As displayed in Figure 2.4.D, the types of features that lead to isomerism in *N*-glycans are:

- Isomeric glycosidic linkage (e.g., α 2-3 or α 2-6 NeuAc)
- Isomeric composition (e.g., monosaccharides GlcNAc or GalNAc, Gal and Man)
- Isomeric branching (e.g., a bisecting GlcNAc or a third antenna-GlcNAc bound to C6- or C3-Man)
- Isomeric anomeric configuration (linkage oriented in α or β anomeric position)

The interpretation of glycoproteomic data (collected via MS) is to some extent simplified by allowing assumptions referred to the biosynthetic pathways and glycosyltransferases (genes) expression that generally participate in the biosynthesis of *N*-glycans in mammalian cells or, even more specifically, in human cells [59, 81]. However, several isomeric glycosidic bonds are displayed in human *N*-glycans, and their relevance is demonstrated by the vast diversity of glycan-binding proteins known as lectins.

Discovered in 1888, lectins bind specific glycan epitopes (glycoepitopes) and are further classified by their carbohydrate-recognition domains [82]. For example, galectins (or S-type lectins) recognize glycans carrying β -galactose, P-type lectins are specific to mannose 6-phosphate containing glycans, and siglecs (I-type lectins) have affinity towards sialylated glycans [82, 83]. The interactions through these glycan-binding molecules establish a complex “communication process” within an organism, where the influence of each *N*-glycan at each *N*-glycosylation site is essential for a biological function [84, 85]. For instance, L-selectin is a transmembrane lectin protein of leukocytes that works as cell-adhesion molecule [86]. Studies estimate the expression of 50,000 to 70,000 L-selectins per leukocyte plasma membrane [87, 88]. The dominant ligand of this lectin is the sulfated sialyl Lewis x structure, present on *O*- and *N*-glycans of some proteins [86]. Different analytical approaches allow researchers to describe *N*-glycan structures that are valuable for industry and medicine [89]. The selection of the analytical approach depends on the targeted aim, sample complexity, and accessibility of the resources required, as it will be discussed in the following section.

2.2. Analysis of *N*-glycosylation of proteins in human blood plasma

The methods applied to characterize the *N*-glycosylation profile of proteins are shown in Figure 2.5 [65]. The study of protein *N*-glycosylation can be performed on three levels: I) glycoprotein, II) *N*-glycopeptide, or III) *N*-glycans.

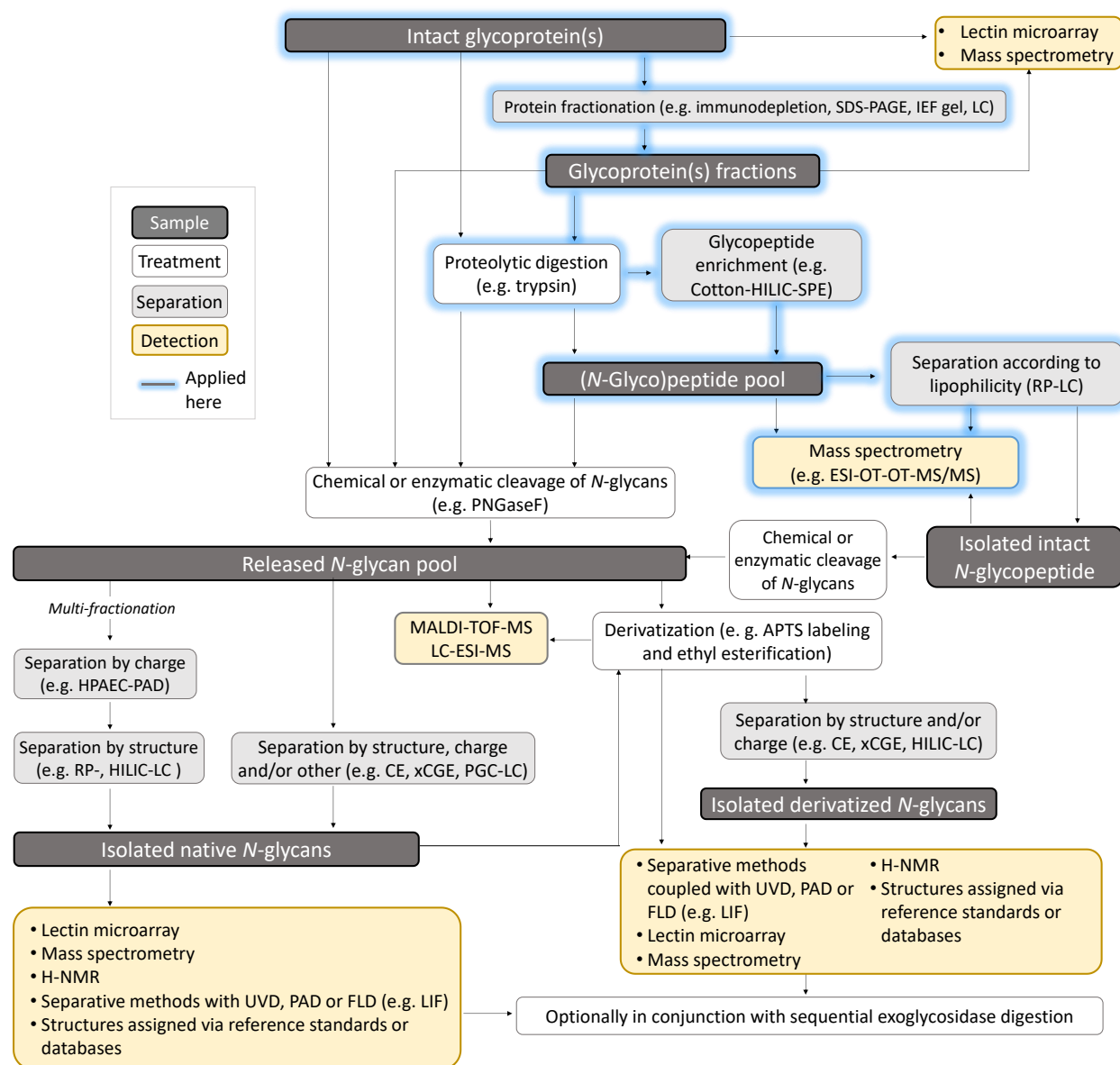


Figure 2.5. Methodologies for the analysis of glycoproteins. Modified from Lottspeich and Engels [65]. SDS-PAGE: sodium dodecyl sulfate-polyacrylamide gel electrophoresis, IEF: Isoelectric focusing, LC: Liquid chromatography, HILIC: Hydrophilic interaction liquid chromatography, SPE: solid-phase extraction, RP: Reversed phase, ESI: Electrospray ionization, OT-OT-MS/MS: orbitrap tandem mass spectrometry, PNGaseF: peptide-N-Glycosidase F, MALDI: matrix-assisted laser desorption/ionization, TOF: time-of-flight, MS: mass spectrometry, HPAEC: High-performance anion-exchange chromatography, PAD: Pulsed amperometric detector, APTS: aminopyrene trisulfonate, CE: capillary electrophoresis, xCGE: multiplexed capillary gel electrophoresis, LIF: laser-induced fluorescence detection, FLD: fluorescence detector, UVD: Ultraviolet light detector, H-NMR: Nuclear magnetic resonance (based on hydrogen atoms).

2.2.1. Approaches for the study of glycoproteins

The analysis of protein *N*-glycosylation includes methods for sample preparation, separation, and detection [90]. On the one hand, the research on intact glycoproteins is typically conducted on isolated pre-characterized proteins (e.g., a monoclonal antibody (mAb) [91]) and shows a distribution of protein glycoforms. Standard methodologies are gel electrophoresis (to evaluate mass or isoelectric point) or a more complex top-down mass spectrometry analysis, for example, to assess the antibody-drug conjugation rate [92]. On the other hand, analysis of released *N*-glycans (glycomic analysis) can provide an in-depth structural description of the *N*-glycan and a reliable relative quantification [90, 91]. A disadvantage of glycomic research is that once *N*-glycans are released, information about their linkage to each *N*-glycosylation site in a protein (micro-heterogeneity) is lost. Compared with these two approaches, only the analyses on intact *N*-glycopeptides (bottom-up *N*-glycoproteomic analysis) can deliver information about the *N*-glycan and the carrier *N*-glycosylation sites on the protein. In a bottom-up analysis, the protein identity is inferred from the peptides detected via MS analysis after a proteolytic digestion [93]. When this analysis is applied to a (glyco-)protein mixture is named shotgun (glyco-)proteomics [93], and both the unmodified peptides and glycopeptides (“(glyco-)peptides”) are preferably characterized via high-performance liquid chromatography with tandem mass spectrometry analysis (LC-MS/MS) [94]. Dedicated MS fragmentation methods, incorporated through two-step MS techniques (MS/MS), can facilitate the interpretation of the composition and branching of *N*-glycans present in intact glycopeptides, with limitations regarding structural description.

To explain the reasoning behind the selection of the methods applied in this thesis, a brief description of blood plasma and the analytical techniques frequently applied to the analysis of protein *N*-glycosylation in this sample will be presented.

2.2.2. Human blood plasma

Human blood is a fluid that connects organs and tissues along the human body, transporting oxygen, nutrients, signaling molecules, etc., as part of its physiological function [95]. As shown in Figure 2.6, blood plasma proteins are distributed in a concentration range of more than ten orders of magnitude [24, 25, 96]. The complexity of the blood plasma protein mixture is high since hundreds of different proteins circulate in the whole concentration range [96]. During its circulation, the blood plasma also collects diverse glycoproteins, including tissue leakage proteins [24, 95]. Found in the low concentration range in blood plasma, the tissue leakage glycoproteins are potential biomarker candidates due to their organ specificity and intracellular microenvironmental changes reflected in the biosynthesis of the *N*-glycans attached [5, 17-23, 73, 96]. However, exploring the low-abundant blood plasma glycoproteins is masked

by the high-abundant proteins whose concentration is several thousand times higher than the tissue leakage proteins (Figure 2.6). Detection of tissue leakage proteins is also hindered by the highly diverse set of blood plasma proteins from the middle- to low-abundance range and their multiple glycoforms [77]. Thus, a sample preparation workflow to access low-abundant blood plasma glycoproteins must include appropriate sample fractionation methods [97]. In the case of site-specific *N*-glycan identification, MS methods and the data analysis workflow must allow the recording and processing of large lists of *N*-glycoproteomic data to cover the complexity of the protein mixture in the middle- to low-abundant range [98]. In the section below (section 2.2.3), a comparison of approaches, used for analyzing the *N*-glycosylation of human blood plasma proteins, is given. Next, methods selected for conducting this thesis are described (section 2.2.4).

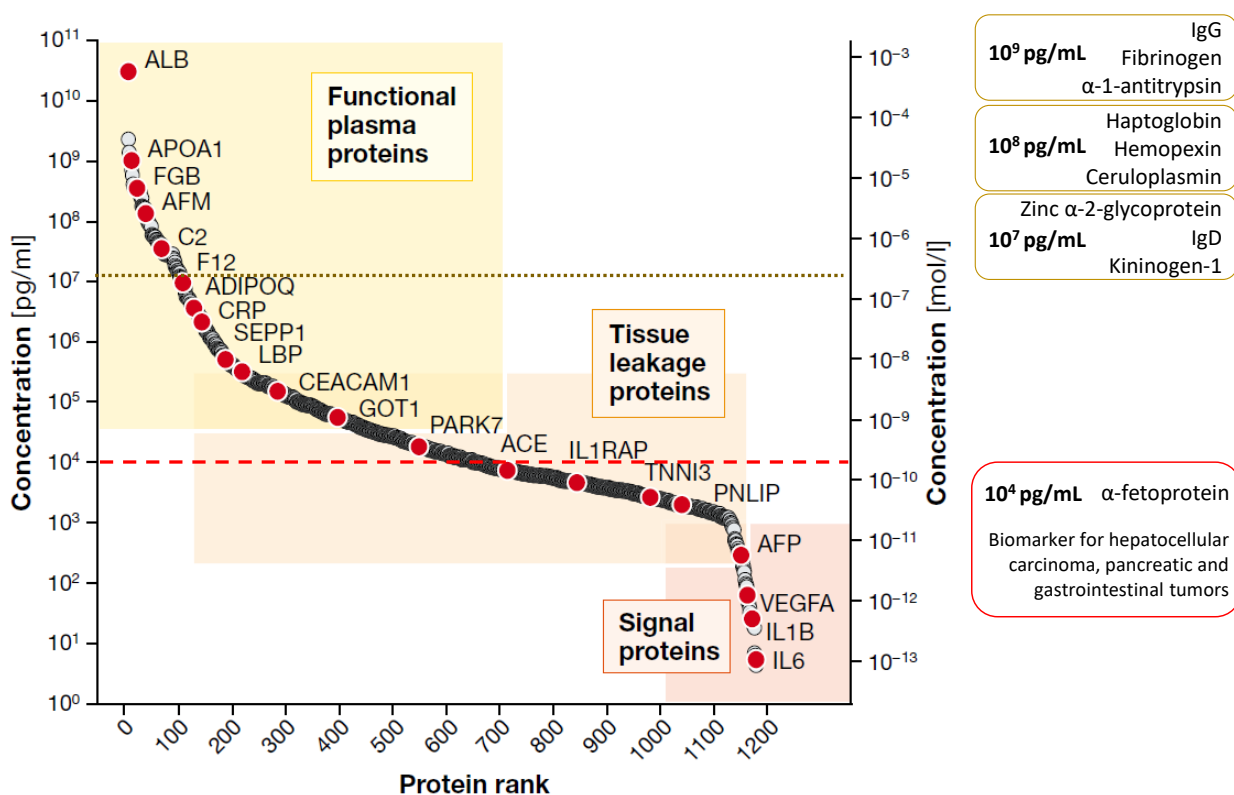


Figure 2.6. Distribution of blood plasma proteins over a concentration range of more than ten orders of magnitude. The brown dotted line indicates the concentration threshold from a few proteins of the acute phase [25]. The red dashed line indicates the approximate concentration of the biomarker α -fetoprotein referred to by Kirwan et al. [5]. Plot modified from Geyer et al. [96] Blood plasma protein concentrations reported by Nanjappa et al. [25].

2.2.3. Standard methods applied for analyzing the *N*-glycosylation of blood plasma proteins

Methodologies for the high-throughput study of blood plasma protein *N*-glycans involve primarily lectin microarray, hydrophilic-interaction ultra-high performance liquid chromatography with fluorescence detection (HILIC-UHPLC-FLD), multiplexed capillary gel electrophoresis with laser-induced fluorescence

detection (xCGE-LIF), and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) [8, 9, 99-101]. These methodologies are included in the MIRAGE guidelines ("Minimum Information Required for A Glycomics Experiment"), which frame the best practices for reporting experiments on released glycans [102-105]. As Figure 2.5 shows, glycomic analyses can be combined with different preparative and separative methods according to the investigative purpose. These methods have been effectively applied to cohort vs control studies. For example, the performance of HILIC-UHPLC-FLD, xCGE-LIF, and MALDI-TOF-MS methods was compared for investigation on protein *N*-glycosylation changes in human blood plasma between women with rheumatoid arthritis and healthy controls [9]. Results suggested that, despite slight differences in the relative quantification of the *N*-glycans given by each method, they agree in most of the *N*-glycosylation changes. For example, all techniques demonstrated that during pregnancy agalactosylated core-fucosylated *N*-glycans decrease while the tri-antennary antenna-fucosylated and the mono-galactosylated core-fucosylated *N*-glycans increase in both healthy and patients. Also, all methods demonstrated that galactosylation increases during pregnancy and decreases after birth in both patients and healthy women. This study also reflected favorable features of each method. For instance, MALDI-TOF-MS combined with sialic acid derivatization (an optional preparative method) showed advantages regarding the characterization of sialylated *N*-glycans and large structures such as tri- or tetra-sialylated [106]. The drawback of the MS techniques is the uncertainty of changes on the *N*-glycan arms (changes on α 1-6 arm or α 1-3 arm) and the problematic identification of glycosidic bonds. In contrast, HILIC-UHPLC-FLD and xCGE-LIF can elucidate these structural features. The annotation of glycosidic bonds in an *N*-glycan structure detected through HILIC-UHPLC-FLD or xCGE-LIF relies on standards or sequential enzymatic degradation [90]. Unanticipated *N*-glycan structures can be explored by combining additional preparative methods, yet it requires more resources and analysis time [45, 106].

Alternatively, lectin microarrays use the affinity of glycoepitopes for a well-defined collection of lectins (glycan-binding proteins) fixed on micro-slides [91]. Compared to the methods mentioned above, lectins can have narrow or broader epitope specificity and sometimes bind in a non-linear range; thus, their qualitative and quantitative results are displayed differently [90]. Nonetheless, lectin microarrays can detect not only *N*-glycans but also *O*-glycans available on the protein surface and provide hints on the total glycome phenotype of the analyzed protein sample. Compared to MS, lectin assays can only recognize glycoepitopes represented in the microarray. Therefore, it is not preferable as an explorative method for identifying rare or unanticipated *N*-glycan structures; in this aspect, MS techniques represent an advantage.

Bottom-up *N*-glycoproteomic analysis on blood plasma samples has also been applied for cohort and case-control studies. The two major mass spectrometric approaches are MALDI-TOF-MS/MS and LC-MS/MS

[94]. In tandem mass spectrometry methods, a precursor ion is directed from a first scan (MS^1) to a second analysis, in which the selected precursor ion is fragmented [91]. Then, the generated fragment ions are detected in a second scan (MS/MS also named MS^2), which facilitates the annotation of the precursor ion. MALDI-TOF- MS/MS is high-throughput and suitable for targeted (or pure) proteins, while LC- MS/MS is usually not high-throughput and suitable for complex protein mixtures [94]. Both methodologies can achieve the site-specific description of the *N*-glycans and the identification of diverse glycoproteins [91, 94]. For reporting on *N*-glycoproteomics experiments, the guidelines for proteomic analysis, published by the HUPO project, must be incorporated in addition to the MIRAGE guidelines [107, 108].

Although both glycomic and glycoproteomic analysis have been applied for biomarker discovery, it has been reported that glycoproteomic analysis can achieve the identification of more specific biomarkers, not only to identify a disease but also the level of damage within a particular organ [5, 109, 110]. Nevertheless, identifying specific biomarker candidates in blood plasma or serum requires expanding the detection of glycoproteins towards the middle to low protein concentration range (in-depth *N*-glycoproteomic analysis). In an early clinical discovery stage, in-depth glycoproteomic analysis is convenient. Subsequently, high-throughput methods are suitable when targeted biomarker candidates have to be evaluated in a broader cohort or case-control population study. To do an in-depth glycoproteomic analysis, available sample preparation strategies integrate dedicated protein fractionation techniques, such as immunodepletion of high-abundant blood plasma proteins, and help to reduce the complexity of the protein mixture [97].

2.2.4. In-depth analysis of the low-abundant *N*-glycoproteome

2.2.4.1. Sample preparation for low-abundant blood plasma proteins

The interest in detecting the low-abundant proteins has driven the development of strategies for conducting in-depth *N*-glycoproteomic analysis. Figure 2.5 displays a path highlighted in blue, that shows the general steps of a workflow for the in-depth bottom-up *N*-glycoproteomic analysis of a complex protein mixture applied hereby. It includes the following steps: I) protein fractionation, II) proteolytic digestion, III) glycopeptide enrichment, IV) sample separation and analysis via LC- MS/MS , and VI) glycoproteomic data evaluation.

Sample preparation workflows for in-depth analysis of blood plasma integrate one or more fractionation method(s) on the protein or peptide level or a combination of both. Typical methods use immunoaffinity-based columns for depleting 2–14 high-abundant proteins (e.g., albumin, IgG, fibrinogen) [30, 31], protein fractionation by molecular weight [31], and fractionation on peptide level by high pH reversed-phase or by size exclusion chromatography [111-113]. Although a sample preparation workflow designed with three

fractionation steps results in a high sensitivity, it is costly and complicated for clinical use [114]. Kaur *et al.* compared eight combinations of fractionation strategies [31]. The first fractionation via low-abundance enrichment proteins or depletion of 6 or 14 high-abundant proteins. Then, a second protein or peptide fractionation via sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) or via reversed-phase chromatography, respectively. The study demonstrated that all the sample preparation workflows tested obtained comparable protein identification numbers. However, strategies using SDS-PAGE combined with high-abundant protein depletion provided data with deeper proteome coverage and preserved quantitative attributes. Furthermore, the study concluded that gel-based approaches are inexpensive, time-saving, and more straightforward.

Peptides are typically generated by proteolytic digestion either in-solution or via filter-aided sample preparation (FASP) [115, 116]. Although both approaches achieve comparable results, FASP can simplify the sample preparation workflow, as it allows the removal of salts or detergents present in the initial protein fraction [117]. Hoffmann *et al.* created an adapted version of the FASP protocol using ammonium bicarbonate as the final enzyme buffer, which enabled eluting the final peptide digest and removing salts by evaporation [118]. Protein digestion is typically performed using trypsin, which cleaves at the carboxyl side of Lys or Arg. This highly specific endopeptidase simplifies the *in-silico* prediction of the peptide sequences generated. Some studies add a second protease (e.g., endoproteinase Lys-C or Glu-C) to reduce the length of some tryptic-digested peptides and improve protein coverage [75, 118, 119]. However, this might increase the cost and time of a sample preparation workflow.

Glycopeptides in the final peptide digest represent a lower share compared to unmodified peptides, causing their detection to be suppressed during LC-MS/MS. Therefore, a glycopeptide enrichment becomes substantial for an *N*-glycoproteomic analysis. Methods for enrichment can be classified into two groups: glycoepitope-binding methods (using lectin, antibodies, aptamers) and physicochemical methods (hydrophilicity, size, charge, or electrostatic interaction) [120]. Glycoepitope-binding methods are suitable for targeted glycans, whereas physicochemical processes can separate a heterogeneous mixture of *N*-glycan structures. Well-established methods implement hydrophilic interaction liquid chromatography (HILIC) using variable stationary phases (e.g. cotton, coated silica microspheres, zwitterionic sulfobetaine bonded phases, and cellulose), and different organic solvents (e.g., acetonitrile, isopropanol, methanol, and ethanol) [121-126]. Although the drawback of HILIC is the co-elution of non-glycosylated peptides, which suppress the signal from the glycopeptides in the enriched fraction, HILIC is preferred in explorative analyses for several reasons [123]. First, glycopeptides bearing a widespread distribution of *N*-glycan structures are unbiasedly enriched. Second, (glyco-)peptides can be recovered without any effect or modification, allowing further analysis. Finally, methods such as cotton hydrophilic interaction liquid chromatography solid-phase extraction (cotton-HILIC-SPE) are high-throughput, accessible, simple, and

compatible with the orthogonal LC-MS/MS analysis [124, 127]. Typically, cotton-HILIC-SPE glycopeptide enrichment include three steps: loading, wash, and elution. The loading fraction has the highest ratio of peptides, and the elution fraction contains the highest ratio of *N*-glycopeptides (more hydrophilic molecules).

2.2.4.2. *Liquid chromatography coupled to electrospray ionization with tandem mass spectrometry for N-glycoproteomics*

A nanoscale liquid chromatography (nanoLC) coupled to electrospray ionization (ESI) with tandem mass spectrometry (LC-ESI-MS/MS) is crucial for maximizing the analysis range and reducing the complexity of the glycopeptide-enriched fraction. The most universally applied stationary phase, in both proteomic and glycoproteomic analysis, is the reversed-phase (RP) stationary phase (e.g., C18 matrix), as the separation depends mainly on the interaction between the peptide moiety and the hydrophobic stationary phase [93, 111-113, 128, 129]. As a result, the separation displays smaller distributions of a peptide sequence holding variable *N*-glycan structures [128]. Other stationary phases such as a HILIC, normal-phase, or a mixture of reverse-phase and porous graphitized carbon (PGC) have been applied in some *N*-glycoproteomic analyses for dedicated purposes [130-132].

Throughout the nanoRP-LC elution time, the coupled MS instrument continuously acquires several spectral collections of two sequential scans, MS^1 and MS^2 [91]. An MS^1 spectrum records several precursor ions (ionized *N*-glycopeptides and other molecules) at a time. Then, the most abundant precursor ions from an MS^1 are sequentially isolated and individually fragmented to generate an MS^2 spectrum. It, therefore, results in a two-step MS, also called the MS/MS technique. The continuous acquisitions between MS^1 and MS^2 scans occur fast; however, this speed is different between the state of the art of the MS instrument. For example, orbitrap Elite-Velos has a scan speed of 4 Hz; orbitrap Q Exactive, 12 Hz; and orbitrap Eclipse, 40 Hz [133, 134]. This significantly impacts the total number of MS^1 and MS^2 scans collected per nanoRP-LC-ESI-MS/MS run [133]. The outcome of these instruments is used as the input data for *N*-glycoproteomic data analysis.

Several features can maximize the identification of the *N*-glycopeptides via MS methods [94, 118]. First, high-resolution mass analyzers (e.g., orbitrap) benefit the interpretation of *N*-glycopeptides due to the high mass accuracy of the detected ions [135]. High mass accuracy is critical to discriminate between fragment ions derived from the glycan and peptide moiety. Second, each fragmentation method yields distributions of fragment ions with diverse *m/z* size, allowing to depict the peptide sequence, and the *N*-glycan structure to some extent [135, 136]. Fragmentation of the ionized intact *N*-glycopeptide can be induced through different MS/MS dissociation methods, such as collision-induced dissociation (CID), higher-energy collision dissociation (HCD), electron transfer dissociation (ETD), or electron-

transfer/higher-energy collision dissociation (EThcD) [136-138]. Third, for collision energy modulated dissociation methods, the fragmentation pattern of the *N*-glycan and peptide moiety depends on the fragmentation kinetic energy applied [135]. For example, lower collisional energies generate larger fragment ions than higher collisional energies [118, 139]. Larger *N*-glycopeptide fragment ions can better depict the branching structures on the *N*-glycan [118, 139]. Stepped collisional energies (NCE $35 \pm 15\%$ or NCE $30 \pm 10\%$) applied to intact *N*-glycopeptide ions effectively dissociate the amino acid sequence and the *N*-linked monosaccharide chain into smaller units [118, 136, 137]. This provides higher certainty on both peptide and *N*-glycan composition.

Mass spectrometric methods for *N*-glycoproteomics focus on generating MS² spectra rich in m/z fragment ions of the peptide and *N*-glycan, as shown in Figure 2.7. The primary ions observed by *N*-glycopeptide fragmentation using CID or HCD methods are the B and Y ions [118, 135, 140]. These fragments derive from cleavages of the glycosidic bonds and not from divisions of the peptide bonds that produce “b and y ions.” This terminology for carbohydrate fragmentation was established by Domon and Costello in 1988 [140]. Also called oxonium ions, the B_i ions fragments present the terminal (non-reducing end) oligosaccharide unit. The opposite Y_j fragment ions contain the reducing oligosaccharide unit attached to the peptide (aglycone)—the subscripts specify the cleavage position relative to the corresponding terminal end [140]. At higher collisional energies (e.g., stepped NCE), cascade fragmentation events occur, resulting in b and y ions but also sugar ring fragmentations (Figure 2.7) [135, 140]. For example, *N*-glycopeptide fragmentation typically generates a Y ion composed of the peptide plus the innermost HexNAc-ring fragment [$M_{\text{peptide}} + H + {}^{0.2}X \text{ HexNAc}$]⁺ [118, 141-143].

Isomeric *N*-glycan structures are the most challenging structural features to discern via MS methods. Technology in this field continuously developed towards a sufficient full structural description of *N*-glycan structures [129]. Recent strategies propose the use of hybrid fragmentation methods (e.g., EThcD), the application of multiple fragmentation kinetic energies on the same precursor ion (e.g., integrating both HCD NCE stepped and HCD NCE 20 fragmentations [118]), or the use of ion mobility as part of fragment ion identification [136, 137, 144]. Demonstrated for blood plasma and other protein samples, MS/MS dissociation methods integrating both HCD NCE stepped and NCE 20 increase the identification and description of intact *N*-glycopeptides as demonstrated by Hoffmann *et al.*, Sanda *et al.*, and Shen *et al.* [118, 139, 145]. Thus, both HCD type of MS/MS dissociation methods were applied in this thesis.

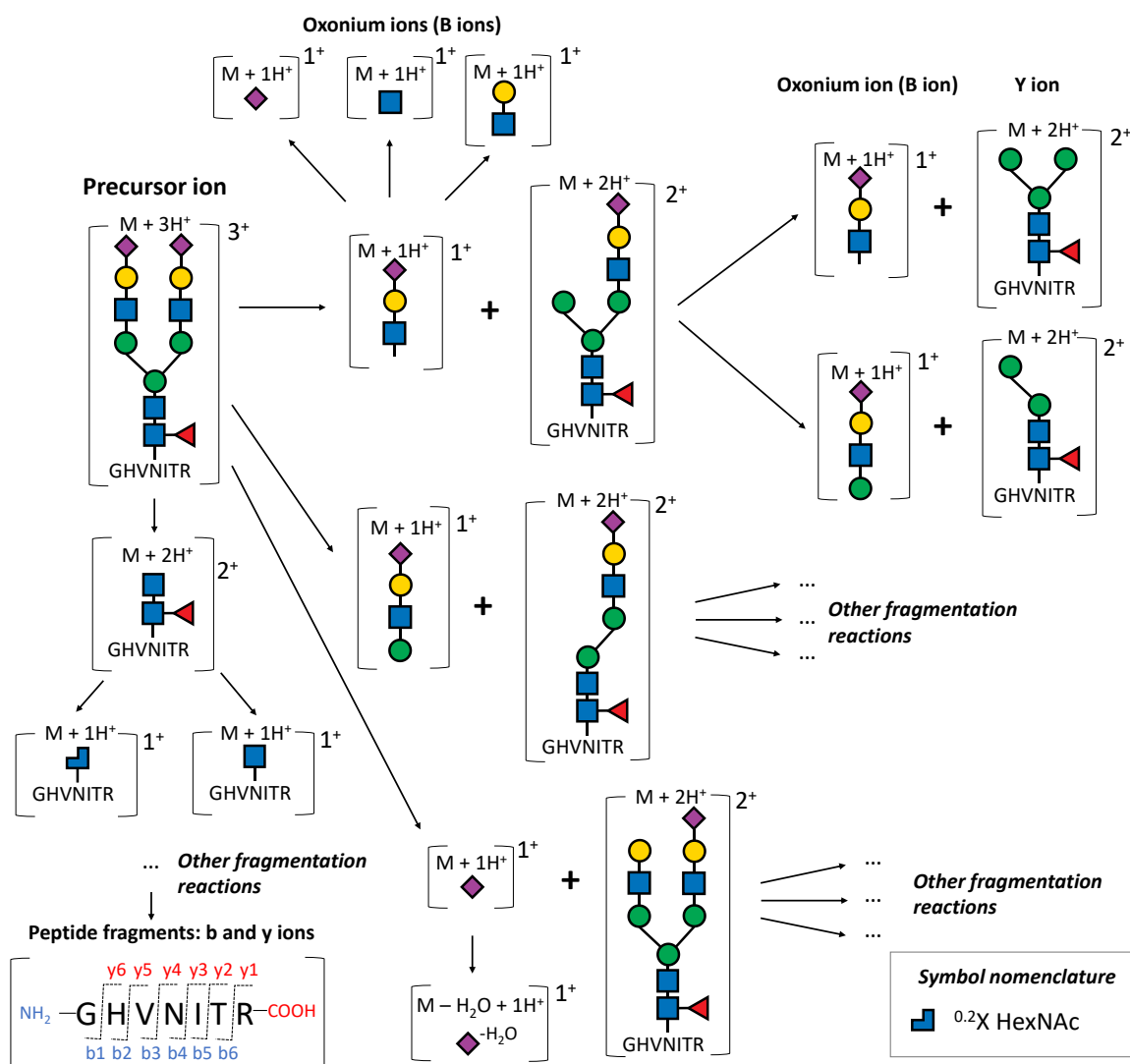


Figure 2.7. Fragmentation of a complex-type *N*-glycopeptide. Primary reactions and selected subsequent fragmentation reactions are shown. Mechanism displayed only with illustrative purposes. Modified from Vekey et al. [135].

2.2.4.3. *N*-glycoproteomic data analysis

An in-depth screening and identification of *N*-glycopeptides aims to understand *N*-glycosylation changes occurring in living organisms. However, the information generated by MS/MS measurements contain thousands of spectra derived and non-derived from *N*-glycopeptides. Hence, an automated screening, matching, selection, and validation of the correct *N*-glycan moiety and carrier peptide must be performed to support the interpretation of the MS/MS data [146]. Input elements required for a software-assisted *N*-glycoproteomic search are:

- Sample preparation parameters, e.g., protease, alkylation of reduced disulfide bonds
- MS instrument and MS/MS method parameters, e.g., mass analyzer, mass tolerance
- Glycan database, e.g., the 182 human *N*-glycans Byonic™ database

- Protein database, e.g., human proteome from UniProtKB
- Peptide modifications, e.g., methionine oxidation, asparagine, or glutamine deamidation
- A high-performance (*N*-glyco-)peptide spectra matching algorithm
- A validation tool

N-glycoproteomic searches aim to “reconstruct” the collection of protein glycoforms of the sample. Therefore, a software-assisted *N*-glycoproteomic search must generate multiple glycopeptide spectrum matches (gPSM) that represent all variable *N*-glycan structures attached to each *N*-glycosylation site of a protein for all glycoproteins present in the sample. A high protein coverage and a reliable identification of each *N*-glycopeptide are necessary to achieve a complete macro- and micro-heterogeneity description of the protein *N*-glycosylation [146]. Due to these challenges, *N*-glycoproteomic analyses were limited to de-glycosylated peptides in the past [120, 146]. Nowadays, tremendous progress has been made in the field, as several software tools enable the identification of intact *N*-glycopeptides [147, 148]. For instance, the software glyXtoolMS, developed by Pioch *et al*, identifies *O*- and *N*-glycopeptides relying on the conserved *N*-glycopeptide fragmentation pattern and oxonium ions [149]. Even so, the “reconstruction” of all protein glycoforms present in a sample remains challenging and is prone to distortion. The relatively low accuracy of glycopeptide spectrum matching software was demonstrated by comparing the outcome of different bioinformatics glycoproteomic analyses of blood plasma and the SARS-CoV-2 spike protein [148, 150]. The comparison showed that a significant fraction of the data resulted in variable (likely misinterpreted) *N*-glycopeptide identifications between software tools, including a considerable variability in the number of total spectra interpreted. In addition, studies suggested that the main drawbacks in achieving accurate results are the scoring algorithms and the settings used for the false discovery rate (FDR) validation methods [148, 150, 151]. This strongly indicates that manual validation, specific-*N*-glycopeptide searches, and searches conditional to selected oxonium ions are still necessary to check for the reliability of gPSM produced by the software [151]. Several of these considerations were implemented in glyXtoolMS software, making it possible to match peptide moieties more precisely, visualize the *O*- or *N*-glycan structure and even discriminate among some isomeric structures, as demonstrated by Hoffmann *et al*. [118, 149]. However, manual validation was still required for MS² spectra that lack information for its correct interpretation, which often occurs with complex samples.

In addition, the accuracy of input elements for glycoproteomic searches influences the identification of the intact *N*-glycopeptides. To maximize the yield of *N*-glycopeptide identifications, it is necessary to investigate the adequate entry parameters, databases, and matching algorithms [148]. For example, although a *N*-glycoproteomic search using the typical database of human blood plasma *N*-glycans would generate several correct *N*-glycopeptide identifications [125, 137, 152], unpredictable changes in protein sequence or *N*-glycan building blocks could be overlooked. In this regard, dedicated glycomic studies have

identified *N*-glycans modified by sulfation, acetylation, or phosphorylation and composed with glucuronic acid in the blood plasma [29, 44, 45, 153-155]. However, these *N*-glycan compositions are frequently not represented in the typical *N*-glycan database entry from *N*-glycoproteomic analysis. Different strategies, such as diagnostic oxonium marker ions, MS² filter layouts, and searches setting a mass range to search for unknown modifications (wildcard search) can assist in identifying rare *N*-glycans [75, 118, 149, 156]. Using a combination of these glycoproteomic strategies, the identification of some rare *N*-glycan structures in blood plasma proteins was carried out in this thesis.

2.3. Role of *N*-glycosylation in normal human physiology and pathophysiology

The analysis of site-specific *N*-glycosylation changes is mainly applicable in two areas. First, in the clinical research the identification of *N*-glycosylation features related to health and disease (so called “carbohydrate-based biomarkers” or its portmanteau “glycobiomarkers”) is required [35, 38-41, 157-161]. Second, in the biopharmaceutical industry, monitoring glycosylation is a critical quality attribute in batch-to-batch production of recombinant proteins [162]. As this thesis focuses on the in-depth *N*-glycoproteomic analysis of human blood plasma, the following section describes some examples of the impact of *N*-glycosylation on health and disease, along with the development of treatment and diagnostic methods.

2.3.1. Normal physiology

The *N*-glycosylation changes during processes in the normal physiological state [3-8]. Aging, hormonal changes, pregnancy, and dietary habits are normal physiological processes that reflect shifts in the *N*-glycosylation profile of human proteins [7, 8, 163, 164]. Knezevic *et al.* evaluated *N*-glycans released from blood plasma protein samples from 1,914 men and women of different ages and pathophysiological states between 2003, 2004, and 2007 [164]. They found significant *N*-glycosylation differences related to gender, age, and hormonal changes, rather than pathologies. For example, reduction of antennary fucosylated *N*-glycans in aging women, while an increment of these structures in men. Due to the lifespan in which this change was observed, this difference was associated with pre- and post-menopausal stages in women's physiology. Posterior studies on aging and *N*-glycosylation changes focused on immunoglobulin G (IgG) [10, 11]. In healthy individuals, the *N*-glycosylation of the high-abundant blood plasma revealed a reproducible correlation curve between the relative abundance of IgG agalactosylated *N*-glycans and chronological age. The levels of these *N*-glycans reduce in middle age and increase again in

old age of life [10, 11]. This shows the benefit of a comprehensive protein site-specific *N*-glycosylation analysis.

Another physiological process with a significant impact on *N*-glycosylation is pregnancy. Bondt *et al.* performed an *N*-glycoproteomic analysis of IgG and IgA purified from the blood plasma of healthy women before, during, and after pregnancy [6, 7, 165]. The results showed that, during pregnancy, there is an increase in levels of bisected *N*-glycans on both IgA *N*-glycosylation sites (N144 and N340) and IgG (N297). Site-specific *N*-glycosylation analysis showed that fucosylation decreased during pregnancy and increased after delivery, uniquely at the N340 site in IgA. While the opposite holds true for triantennary *N*-glycans. Furthermore, sialylation increases during pregnancy on both immunoglobulins but only on the N144 site in IgA. This site is sometimes considered an analog to the IgG *N*-glycosylation site N297, an effector function modulator.

The role of *N*-glycosylation on the effector function of IgG has been exhaustively investigated. First, core-fucosylation diminished binding affinity for the low affinity immunoglobulin gamma Fc receptor IIIa (FcγRIIIa) and lowered the antibody-dependent cellular cytotoxicity [166]. Second, terminal galactosylated *N*-glycans showed an anti-inflammatory effect on women with rheumatoid arthritis during pregnancy [6]. Galactosylation was recently demonstrated to increase binding affinity for the first complement component (C1q) [167]. C1q is a protein complex involved in several mechanisms in the immune system, such as the classical complement pathway activation [168], immune cells function regulation, and recognizing self and non-self-ligands [169]. Third, agalactosylated, asialylated, and afucosylated *N*-glycans of antibodies specific to the HIV virus are associated with reduced virus replication [170]. Fourth, sialylation increased the anti-inflammatory activity of antibodies, although the Fc-mechanisms are not well understood so far [171].

2.3.2. Human immunoglobulin A: a potential therapeutic glycoprotein

Human IgA, investigated in this thesis (chapter 5), has enormous therapeutic potential due to its antiviral potency, mechanisms for eliminating tumor cells, and ability to modulate inflammatory and autoimmune diseases [172-174]. It performs a significant function in the immune mechanisms against mucosal infections, microbiota regulation, and neonatal immunization (Figure 2.8) [175-177]. IgA is the primary secretory immunoglobulin in the mucosa and the second most abundant immunoglobulin in the serum [178, 179]. There are two IgA subclasses in humans, IgA1 and IgA2, which (IgA1:IgA2) ratio in serum is 89:11 [178]. Both subclasses have variants with amino acid substitutions (e.g., IgA2 M319L related to UniProtKB P01877 and PODOX2). In addition, IgA2 is expressed in three genetically controlled polymorphisms, the allotypes m1, m2, and n [180-182]. Compared to IgA2, the amino acid sequence of

IgA1 has an *O*-glycosylated hinge region that lengthens the sequence by 13 amino acids (UniProtKB P01876) [178]. Structurally, IgA1 and IgA2 can be arranged as dimers or monomers. IgA secreted into mucosal tissues (intestinal, nasal, and oral tissues) is dimerized using a joining-chain (J-chain) [178]. Then, it is carried through the epithelial barriers by binding to the polymeric immunoglobulin receptor (pIgR) – this receptor is the precursor of the secretory component required to form secretory IgA [178]. In mucosal tissues, 87% to 97% of IgA is present as a dimer [179]. In contrast, in the blood only 1% to 15% of the IgA exists as dimer [179]. Compared to IgG, the IgA subclasses are richer in *N*-glycosylation sites in the Fc and the Fab region. The two *N*-glycosylation sites in IgA1 are homologous to the ones in IgA2 Fc region, N144-IgA1/N131-IgA2 (on the C_H2 domain) and N340-IgA1/N327-IgA2_{m1/n} (on the C-terminal tailpiece at C_H3 domain) – the IgA2_{m2} shares no homology at the tailpiece sequence [178, 182]. IgA2 has an additional *N*-glycosylation site on the C_H2 domain (N205) and a fourth site, N47, at the C_H1 Fab region. A fifth *N*-glycosylation site (N92) exists only in the IgA2 allotypes m2 and n on the Fab region [178, 182].

IgA displays *N*-glycan remodeling as an effect of physiological and pathophysiological processes [165, 183]. In contrast to IgG, the most abundant serum IgA *N*-glycans in healthy donors are the mono- or di-sialylated bisected and non-bisected di-antennary complex-type *N*-glycans [7, 110, 152, 184]. Core-fucosylated sialylated complex-type *N*-glycans are characteristic at the tailpiece *N*-glycosylation site and not at the C_H2 site (N144-IgA1/N131-IgA2) [7, 110, 152]. Studies reveal that the *N*-glycosylation profile of IgA2 has a higher abundance of hybrid- and oligomannose-type *N*-glycans than IgA1 [152, 184]. A brief overlook of the relationship between IgA effector functions and IgA *N*-glycosylation can be observed in Figure 2.8. A more comprehensive description is given in chapter 5.

The presence of sulfated *N*-glycans in IgA and other glycoproteins has been overlooked due to its challenging detection via established glycomic and glycoproteomic methods [185]. Recently, the detection of sulfated *N*-glycans via glycomic analyses was reported; however, their location within the many IgA *N*-glycosylation sites was not described [28, 44, 45, 154, 186, 187]. This was a reason to include in this thesis corresponding investigations for the site-specific identification of IgA rare *N*-glycans. Chapter 5 will focus on applying the developed in-depth *N*-glycoproteomic workflow to the site-specific *N*-glycosylation analysis of IgA. With this, sulfated and other rare *N*-glycans could be considered when profiling human IgA *N*-glycosylation.

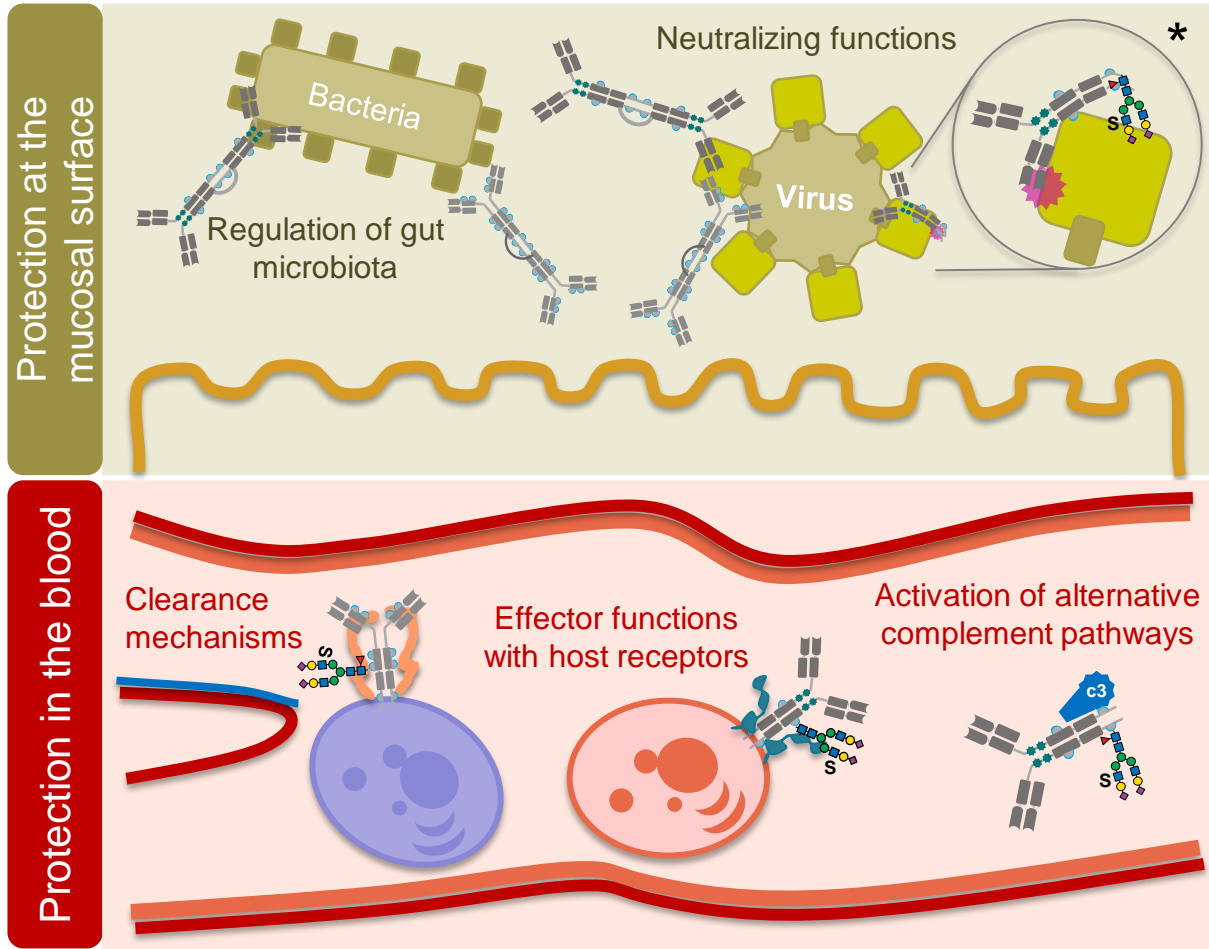


Figure 2.8. IgA functions in the immune system influenced by N-glycosylation. In the mucosa IgA N-glycosylation modulates dimer assembly in IgA1 [188], its critical for binding to complement C3 protein[188], in addition sialylation enhances its anti-viral neutralizing activity by binding to sialic-acid binding viruses (indicated with an asterisk *) [172]. In the blood IgA N-glycosylation affects the serum half-life of IgA1[189]. Specifically, N-glycosylation at IgA1 Fc domains modulates the anti-inflammatory activity [174, 184], and IgA2 N-glycosylation influences its clearance from blood[189]. Glycosylation with HexNAc sulfation is shown only as an example of a future study aim.

2.3.3. Pathophysiology

The investigation of N-glycosylation changes concerning pathophysiological states evolved by applying glycomic and glycoproteomic analytical techniques on longitudinal studies and model organisms [100]. CDG, autoimmune diseases, cancer, and neurodegenerative disorders are pathologies associated with alterations of the protein N-glycosylation [190]. The “global burden of disease study” shows a dramatic increase in the incidence rate of pathologies over the last 20 years [191-193]. Figure 2.9 displays the global expansion of neurodegenerative disorders such as Alzheimer’s disease.

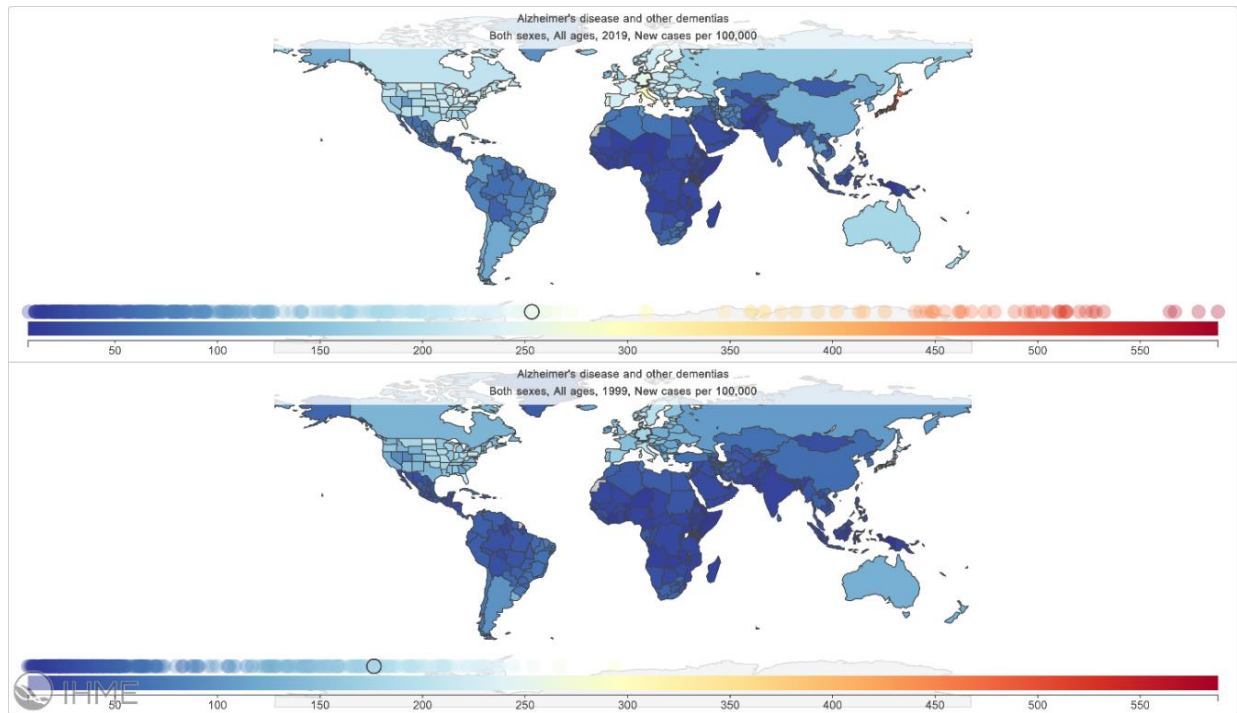


Figure 2.9. Comparison of incidence rate in 1999 versus 2019 for Alzheimer's disease and other dementias as cause of death or injury. All ages and both sexes are included. The circle indicates the number of new cases in Germany. Taken from the Institute for Health Metrics and Evaluation (IHME) [192].

The sixth cause of death in the world, the “tracheal, bronchus, and lung cancer,” is also raising [194]. A more particular case is the number of new patients with pancreatic cancer. As the map in Figure 2.10 shows, the incidence rate in Germany and Finland, from 1999 to 2019, was more prominent than in other countries – from 18.12 and 17.01 per 100,000 in 1999 to 27.05 and 26.86 per 100,000 during 2019 in Germany and Finland, respectively [195].

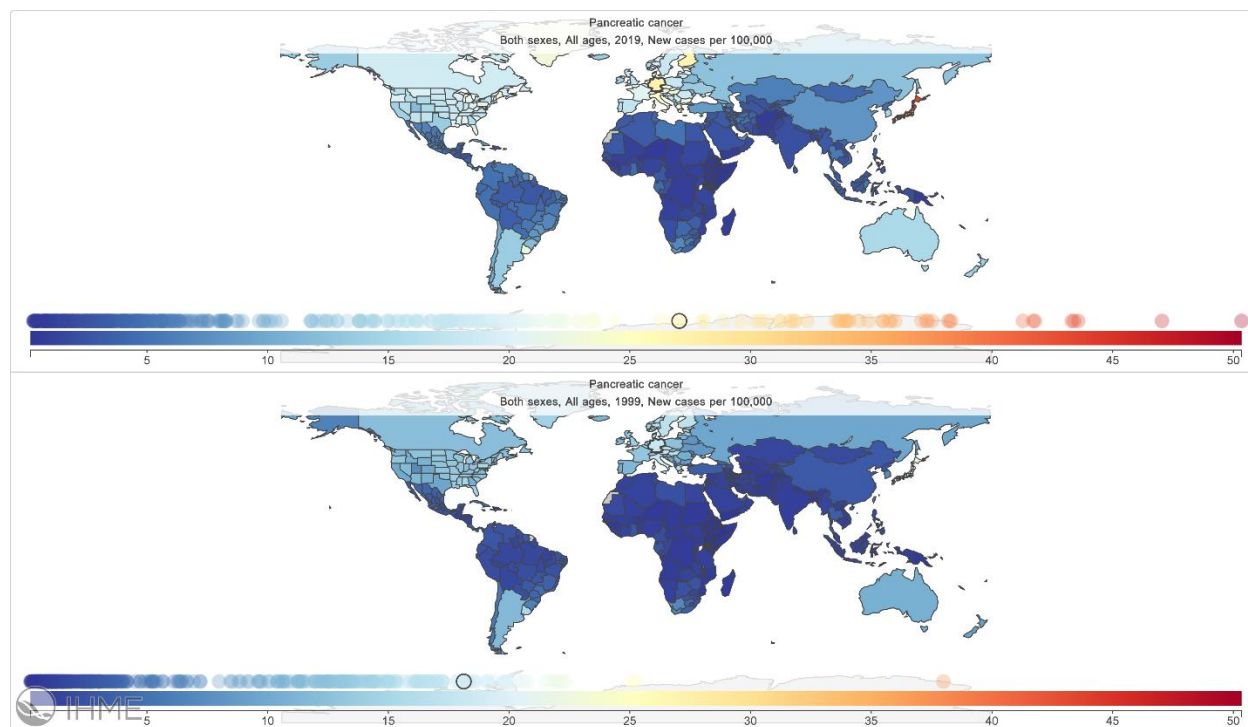


Figure 2.10. Comparison of incidence rate in 1999 versus 2019 for pancreatic cancer as a cause of death or injury. All ages and both sexes are included. The circle indicates the number of new cases in Germany. Taken from the IHME [195].

The rising risk of developing these non-communicable diseases and the low quality of life during disease progression foster the research on biomarkers for early diagnostic and drugs or therapeutic methods. Non-invasive and accessible diagnostic procedures with appropriate sensitivity and specificity at early disease stages are rare, even for frequent diseases such as cancer, Alzheimer's disease, and autoimmune diseases [196-198]. In addition, the method developed in this thesis can contribute to identify protein site-specific *N*-glycosylation features within the low abundant *N*-glycoproteome, which could be tested as potential biomarkers helpful in clinical decision-making or could support the design of biopharmaceuticals.

2.3.3.1. Autoimmune diseases

An autoimmune disease arises when the immune system loses the ability to differentiate between self and non-self (or foreign) antigen [199]. This ability is known as self-tolerance, and it is integrated by the central tolerance (clearance of autoreactive lymphocytes) and peripheral tolerance (suppression of non-cleared autoreactive lymphocytes released to tissues) [199]. These disorders are heterogeneous as they can affect any self-antigen (cell, tissue, organ, or system). Their causes are multiple, e.g., genetic, epigenetic, hormonal, exposure to chemicals or infections, etc. [199]. The worldwide comparison of the incidence of an autoimmune disease such as rheumatoid arthritis is shown in Figure 2.11. The map displays an increment in the frequency of new cases between 1999 and 2019, especially in western countries [193]. Diagnosis is complex and cannot be accomplished with a single test; for example, the AIDs diagnostic yield

employing gene sequencing techniques is 15% using ten genes and 23% using 502 genes[200]. Several tests are necessary for diagnosis, such as immunoglobulin and complement serum protein levels, serological screening for antibodies against self-antigens, and, to measure disease activity, levels of inflammatory markers (e.g., erythrocyte sedimentation rate and C-reactive protein) [201]. Thus, the patient experiences a long diagnostic process while the disease continues to worsen [198]. Even after a successful diagnosis, the path toward treatment is uncertain.

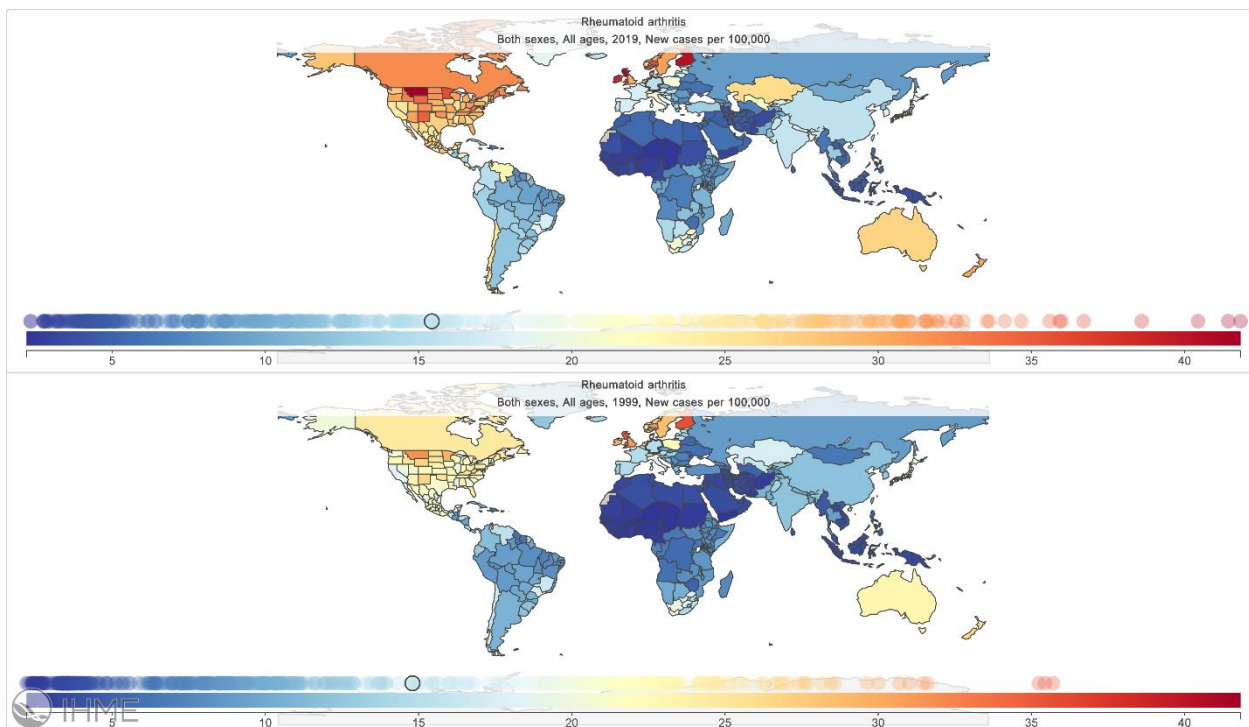


Figure 2.11. Comparison of incidence rate in 1999 versus 2019 for rheumatoid arthritis as cause of death or injury. All ages and both sexes are included. The circle indicates the number of new cases in Germany. Taken from the IHME[193].

Research is progressing to comprehend the onset and modulation of AIDs. The glycoanalytic field has demonstrated that aberrant *N*-glycosylation at the IgG Fc region is associated with different autoimmune diseases, such as rheumatoid arthritis, ulcerative colitis, multiple sclerosis, systemic lupus erythematosus, and autoimmune thyroid disease [6, 35, 38-41, 157-159, 202]. Figure 2.12 displays the alterations of *N*-glycosylation observed for each disease compared to healthy control groups [36, 37]. As explained before (section 2.3.1.), the effector functions of IgG are modulated by *N*-glycosylation [6, 166-171]. Rheumatoid arthritis is probably one of the most studied AIDs in terms of the influence of alterations in IgG *N*-glycosylation and disease activity. Early studies show a significant decrease in galactosylated *N*-glycans of serum IgG correlating with higher disease activity symptoms [11]. Later, it was demonstrated that a reduction in galactosylated *N*-glycans only from the anti-citrullinated protein antibodies (ACPA) starts one year before diagnosis of rheumatoid arthritis [76]. The ACPA are autoantibodies (mainly IgG1)

commonly detected in patients with rheumatoid arthritis [76]. More studies are needed to clarify the causality of the disease. However, the early-altered ACPA-IgG1 *N*-glycosylation suggests an intracellular microenvironmental change that may play a role in the development of systemic inflammation.

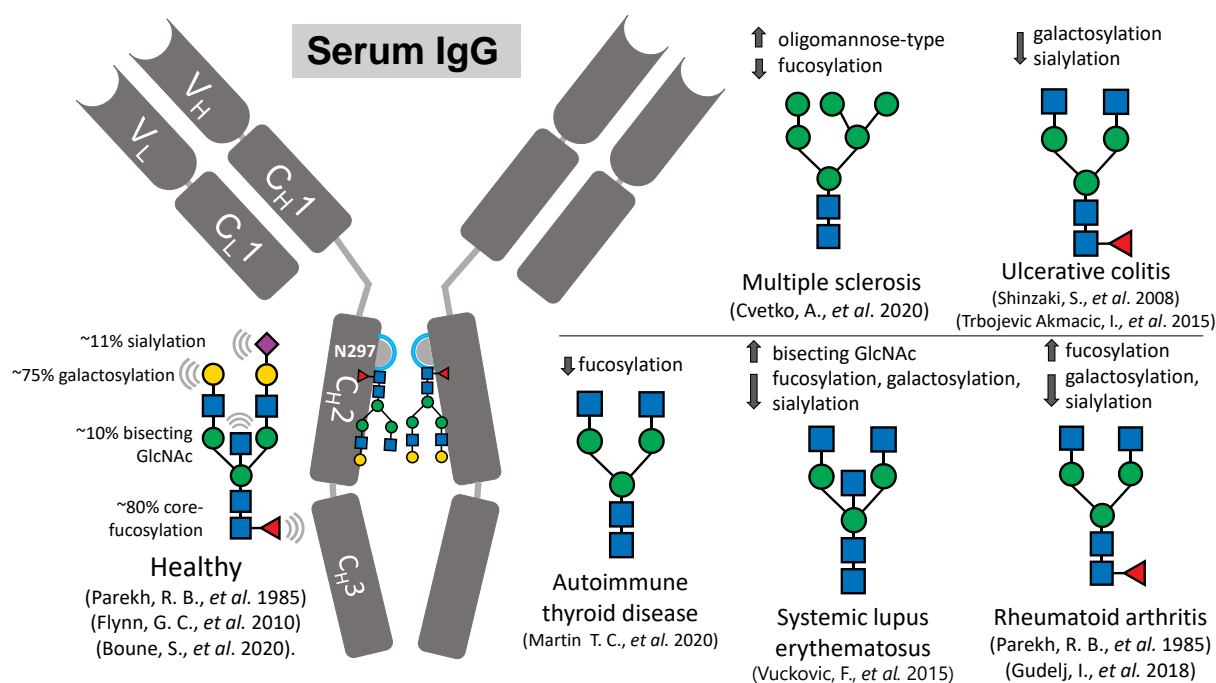


Figure 2.12. Examples about *N*-glycans displayed in human serum IgG in different AIDs. Arrows indicate a higher or lower value relative to the healthy levels of the corresponding *N*-glycosylation feature [35-41]. Adapted from Zabczynska *et al.* [157].

Moreover, the pronounced incidence of rheumatoid arthritis cases during menopause was associated with a decreased estrogen content in serum [203]. The study demonstrated that estrogen levels were related to the abundance of galactosylated and sialylated serum IgG *N*-glycans and that estrogen therapy increased the quantity of sialylated *N*-glycans in total serum IgG. This therapy reduced disease activity. The anti-inflammatory effect of sialylated IgG *N*-glycans has been used by replacing their immunoglobulins with intravenous immunoglobulins (IVIg) purified from healthy donors, as a therapeutic strategy in patients with other AIDs [204]. However, the production of IVIg is donor-dependent, and the demand is often higher than its production capacity. Other strategies are being developed to increase immunoglobulins' production, such as recombinant production of sialylated monoclonal antibodies, controlling the glycosylation profile required [37].

Cell surface *N*-glycans and the *N*-glycan receptors are not only involved in modulating cell proliferation, cell differentiation, and pathogen destruction mechanisms but also in recognition mechanisms involving self and non-self antigens [84, 205]. For example, the abundance of branched *N*-glycans on the CD25

protein positively influences T cell's immune tolerance [206]. Changes in the *N*-glycosylation of cell-surface receptors on immune cells occur during AIDs in damaged tissue or infection [205]. For instance, in inflamed rheumatic joints, the synovial fibroblast cells are observed to gain an aggressive pro-inflammatory phenotype [202, 207]. This phenotype is described as the cell-surface *N*-glycans displaying less sialylated *N*-glycans; the level of sialylation correlates with disease stages in rheumatoid arthritis or remission [207]. Synovial fibroblast cells support the nutrition and structure of joints in a normal physiological state [202]. Shifts in *N*-glycosylation of the plasma cells' surface proteins can upregulate pro-inflammatory reactions, which are especially harmful during autoimmune diseases [205]. This demonstrates the critical role of research about alterations of *N*-glycosylation and encourages including glycobiology in human medicine and the relevance of its analysis.

2.3.3.2. Congenital disorders of glycosylation (CDG)

CDG are innate abnormalities of metabolism caused by genetic defects affecting the expression of proteins involved in the glycosylation pathway [208]. The number of congenital disabilities has decreased in the last 20 years [209]. However, the specific share of CDG worldwide needs to be better researched. The prevalence of CDG in Europe was 0.1-0.5 per 100,000 by 2018 [210]. By October 2021, worldwide scientific publications about CDG patients showed that the typical first diagnostic method is isoelectric focusing either of serum transferrin or serum apolipoprotein C-III [208]. This method enables the diagnostic of a CDG type I or type II. A transferrin type 1 pattern (reduction of tetra-sialotransferrin and augment of di- and asialotransferrin) indicates a CDG- I, genetic defects affecting the glycan transference to the aglycone. A type 2 pattern (increase of tri- and mono- sialotransferrin) suggests CDG type II, disabilities affecting glycan remodeling [208, 210]. Yet, it mainly detects *N*-glycosylation defects related to the absence of *N*-glycosylation or sialylation, not *O*-glycosylation or fucosylation defects. Although, isoelectric focusing methods using serum apolipoprotein C-III were applied for identification of *O*-glycosylation defects, glycosylation traits are only barely characterized [210]. To obtain accurate information for diagnosing a CDG type I or type II, it is necessary to integrate both glycomic/glycoproteomic analysis and exome/genome sequencing, as demonstrated by Rapp and co-workers [32-34]. The most frequent CDGs are PMM2-CDG, FKTN-CDG, EXT1/EXT2, ALG6-CDG, and PIGA-CDG [208]. The research around the diverse CDG conditions has improved our understanding of glycosylation pathways [20, 211]. Glycobiology studies carried out by multidisciplinary teams have also identified various glycosylation deficiencies that can be overcome by including an oral supplement or by gene therapy and improve the life quality of the individual [3, 212, 213].

2.3.3.3. Cancer

Tumor tissues are initially a clustered abnormal cell growth [214]. This become malign when their excessive proliferation invades and damages other tissues until tumor cells are released into the lymphatic or circulatory systems and initiate a second tumor (metastasis) [214]. Current cancer diagnostics via non-invasive methods includes positron emission tomography (PET) and X-ray. However, imaging analysis relies on screening of the tumor size with dimension is very small at early stages of cancer [215]. In contrast, omics analyses conducted on samples obtained by liquid biopsy (e.g., blood plasma) have been demonstrated to achieve the high sensitivity needed for an early diagnosis [196]. Over the last years, the FDA has approved a few tests using different omics, some based on protein biomarkers [196].

During tumor progression, the observed tissue-associated *N*-glycans show an increment of sialylation (especially α 2-6 sialic acid), core-fucosylation, and branching [216-218]. These patterns are related to the activation of cell growth mechanisms for tumor progression and development, and are often named “oncofetal” [219, 220]. Glycomic or glycoproteomic analysis shows promising molecular candidates [221, 222]. Kontro *et al.* compared pancreatic cancer patients, acute pancreatitis vs healthy controls, identifying changes in the abundance of sialylated *N*-glycopeptides from acute-phase proteins (e.g., IgG, transferrin, haptoglobin) [222]. However, the clinical experience in proteomics and glycomics or glycoproteomics based on acute-phase proteins have shown a high risk of obtaining false positives (low specificity) [223]. Acute-phase proteins, such as IgG, transferrin, and haptoglobin, are commonly high-abundant proteins in blood plasma [25]. The precise identification of a tissue-specific disease can improve with the low-abundant glycoproteins. This was observed with the serological α -fetoprotein, whose fucosylation level allows the detection of hepatocellular carcinoma [5, 13, 42]. In the blood serum of patients with hepatocellular carcinoma, α -fetoprotein has a concentration of approximately 20 ng/mL [5]. This demonstrates the potential of in-depth *N*-glycoproteomic analysis for biomarker discovery.

2.3.3.4. Alzheimer's disease (AD)

AD starts with a combination of pathogenic factors in the brain, including neurotransmitter disruption, accumulation of toxic aggregates of amyloid- β ($\alpha\beta$), oxidative stress, neuroinflammation, and other aspects [224]. Interestingly, the disease starts many years before the individual manifests mild cognitive impairment [197]. Neuropsychological tests allow a diagnostic up to eight years before disease development [197]. Nonetheless, the study of alterations of *N*-glycosylation in AD can contribute to the understanding of pathological molecular mechanisms, for example, regarding predictors of disease worsening. A significant alteration of glycosylation in the brain with AD was shown by glycans holding the glycoepitope HNK-1 [225]. This is a trisaccharide typically sulfated and described as SO₄-3GlcA β 1-3Gal β 1-

4GlcNAc. The study from Garcia-Ayllon *et al.* implemented an antiHNK-1 monoclonal antibody to evaluate the level of glycoproteins bearing glycans with HNK-1 [225]. A significant reduction of HNK-1-carrying glycoproteins in AD patients was detected. In this study, it remained unclear if this glycan structure belongs to *O*-glycans, *N*-glycans, or both. Nonetheless, this study presumed that the level of glycoproteins carrying HNK-1 is potentially related to the abundance of β -amyloid glycoproteins and not to the expression of glycosyltransferases involved in the synthesis of this sulfated glucuronidated trisaccharide.

N-glycan structures holding HNK-1 glycoepitope are rarely identified in blood plasma due to their minor abundance [29, 153]. Thus, in-depth *N*-glycoproteomic analysis can support the identification of these rare *N*-glycans in blood plasma proteins. Furthermore, *N*-glycoproteomic methods that detect *N*-glycans carrying HNK-1 in blood plasma glycoproteins can be assessed as a non-invasive tool for monitoring the onset and progression of neurological diseases such as AD.

Intentionally blank page

3 Materials and Methods

Information on the experimental methods applied for studies described in chapters 4 and 5 are presented here. Please note that parts of this section are taken from the original publications Zuniga-Banuelos *et al.* 2023, 2025a and Zuniga-Banuelos *et al.* 2024, 2025b [226-229].

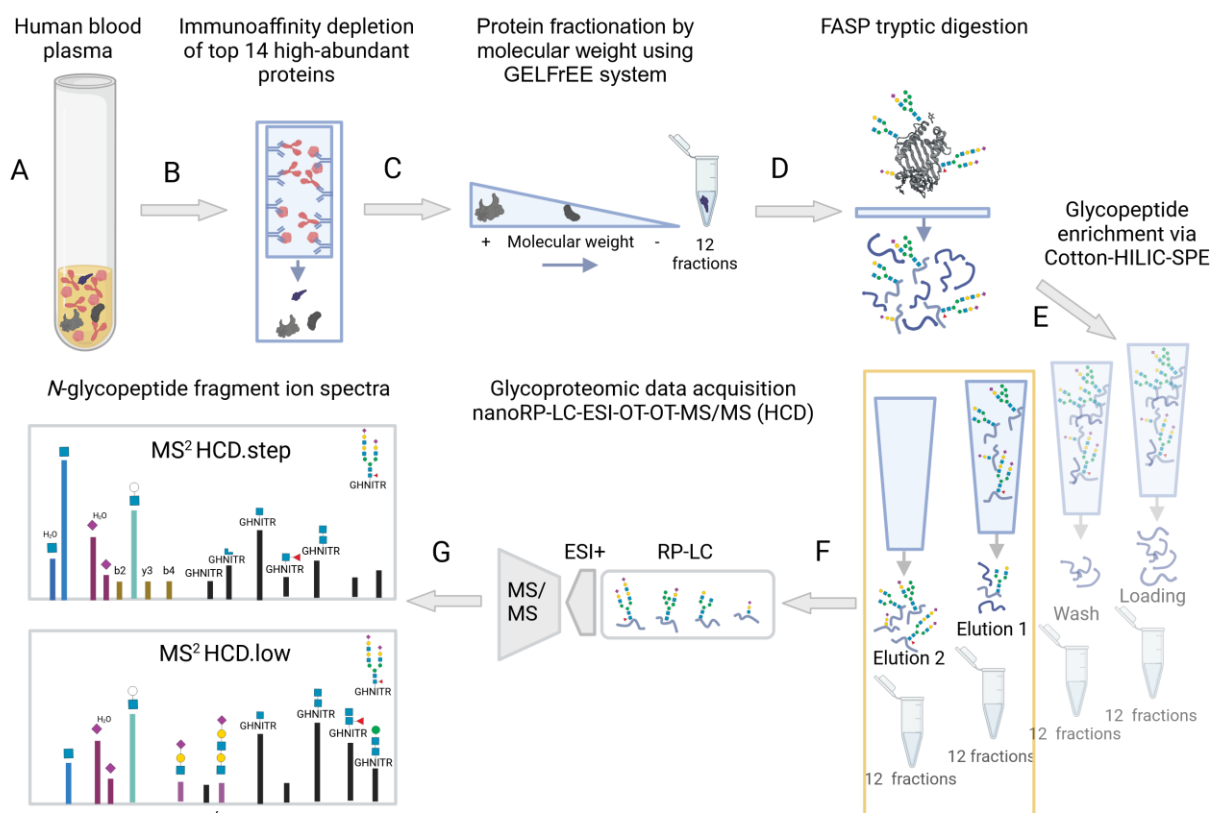


Figure 3.1. Sample preparation workflow and measurement. (A) Blood plasma/serum sample. (B) Immunoaffinity depletion of top 14 high-abundant blood plasma proteins. (C) Separation of blood plasma-depleted fraction by molecular weight through the GELFrEE® system. Collection of 12 fractions. (D) Tryptic digestion via filter-aided sample preparation (FASP). (E) Glycopeptide enrichment via cotton hydrophilic interaction liquid chromatography solid-phase extraction (HILIC-SPE). Collection of four HILIC fractions per protein fraction (loading, wash, elution 1 and 2). (F) Data acquisition through reverse-phase liquid chromatography electrospray ionization tandem mass spectrometry using orbitrap for precursor and fragment ions. (G) Two MS² analyses of each glycopeptide-enriched fraction: higher-energy collision dissociation (HCD) step and low. Image created with BioRender.com

3.1. Reagents, human blood plasma, and IgA samples

The reagents applied were MS grade or the highest purity available. Milli-Q water suitable for LC-MS analysis was freshly obtained from a Millipore Milli-Q® Advantage A10 system (18,2 MΩ×cm, <5 ppb) equipped with a LC-Pak® Polisher filter unit (#LCPAK0001) purchased from Merck Millipore (Darmstadt, Germany). Acetonitrile (ACN, #A955-212) was purchased from Fisher Scientific (Schwerte, Germany). Ammonium bicarbonate (ABC, #09830), 2,2,2-trifluoroethanol (TFE, #808259), potassium chloride (KCl, #104935), potassium phosphate dibasic (K₂HPO₄, #104873), sodium phosphate dibasic (Na₂HPO₄, #106585), formic acid (FA, #56302), DL-dithiothreitol (DTT, #D5545), iodoacetamide (IAA, #I1149), calcium chloride (CaCl₂, #A4689), and sodium dodecyl sulfate (SDS, #75746) were purchased from Merck (Darmstadt, Germany). Sodium chloride (NaCl, #P029.3) was purchased from Carl Roth (Karlsruhe, Germany). Trifluoroacetic acid (TFA, #28904) was purchased from Fisher Scientific (Schwerte, Germany). Sequencing grade modified trypsin was purchased from Promega (#V5111, Madison, WI, USA).

A normal human blood plasma (HBP) sample was used to develop and evaluate the in-depth *N*-glycoproteomic workflow presented in chapter 4. Acquired from Affinity Biologicals, the HBP sample is a pool of normal citrated plasma of at least 20 donors (VisuCon™-F Frozen normal control plasma, FRNCP0105, Ancaster, ON, Canada).

Human IgA purified from normal blood serum was analyzed in chapter 5. Two commercial samples of IgA were purchased from two suppliers. The supplier 1 was Sigma-Aldrich (14036-1MG, St. Louis, MO, USA), and supplier 2 was Athens Research & Technology (16-16-090701, Athens, Georgia, USA).

3.2. Workflow for in-depth *N*-glycoproteomic analysis

3.2.1. Depletion of the top-14 high-abundant blood plasma proteins

The steps applied in the sample preparation workflow are depicted in Figure 3.1. The top 14 high-abundant blood plasma proteins (HAP) were depleted using single-use High Select™ Top 14 Abundant Protein Depletion Midi Spin Columns (#A36371, Thermo Fisher Scientific, IL, USA). These top proteins included albumin, IgA, IgM, IgD, IgE, IgG (kappa and lambda light chains), α-1-antitrypsin, α-1-acid glycoprotein, α-2-macroglobulin, apolipoprotein A1, fibrinogen, haptoglobin, and transferrin. Four immunoaffinity columns were equilibrated at room temperature by end-over-end agitation at 13 rpm (MultiBio RS-24, BIOSAN, Riga, Latvia) 1 h before sample application. Each column was loaded with 70 μL of untreated blood plasma. The columns were incubated for 15 min with end-over-end agitation at 13 rpm. The bottom

of the column was opened to collect the flow-through by centrifugation at $1,000 \times g$ for 1 min (Heraeus Multifuge X 1R centrifuge, Thermo Fisher Scientific, Osterode, Germany). This instrument was used for all centrifugation steps. The slurry was resuspended by adding 1 mL PBS (8 g/L $\text{NaCl}_{(\text{aq})}$, 0.2 g/L $\text{KCl}_{(\text{aq})}$, 0.2 g/L $\text{K}_2\text{HPO}_{4(\text{aq})}$, 1.15 g/L $\text{Na}_2\text{HPO}_{4(\text{aq})}$ pH adjusted to 7.4). The flow-through was recovered in the same tube by centrifuging at $1,000 \times g$ for 1 min. The flow-through from four columns was pooled and referred to as “top 14-HAP depleted sample”. To avoid protein precipitation during the subsequent desalting process, SDS detergent to a final concentration of 0.01% (v/v) was added to the top 14-HAP depleted sample. This protein sample was concentrated ten times using a 3 kDa Amicon Ultra 15 mL device (#UFC900396, Merck, County Cork, Ireland) at $3,000 \times g$ 80 min. The initial concentration of NaCl in PBS (130 mM) was reduced to a concentration below 0.7 mM NaCl by two centrifugation steps of $3,000 \times g$ for 60 min, adding 10 mL water and one last centrifugation step of $3,000 \times g$ for 30 min, adding 3 mL water. The sample volume was reduced to approximately 400 μL using a rotational vacuum concentrator (RVC 2-33 CO plus, Christ, Osterode am Harz, Germany) at 1°C , 0.1 mbar, approximately 2 h. This instrument was used for all drying steps within the sample preparation workflow. The protein sample was quantified by a Pierce™ BCA Protein Assay Kit (#23225, Thermo Fisher Scientific, IL, USA) and stored at -80°C .

3.2.2. GELFrEE® system fractionation

3.2.2.1. Depleted blood plasma fractionation

Protein fractionation by molecular weight was conducted through an eight-channel GELFrEE® 8100 Fractionation System (GELFrEE® 8100 Fractionation System, Abcam, Cambridge, UK). The cartridge (8% Cartridge Kit #42103, Abcam, Cambridge, UK) was prepared according to the manufacturer's instructions. Once desalted and concentrated, an amount of 800 μg of protein from the “top 14-HAP depleted sample” was taken for sample preparation. The taken volume was adjusted to 448 μL with water. Then, GELFrEE® acetate sample buffer (120 μL) and 32 μL of 1 M $\text{DTT}_{(\text{aq})}$ were added. The sample mixture was heated at 50°C for 10 min. The total sample volume was divided by four, loading 200 μg protein sample in each cartridge channel (150 μL). The electrophoresis method was programmed to generate 12 fractions. The electrophoresis loading step started at 50 V and continued for 16 min. Then, 12 fractions were collected in the following 152.5 min. The amount of protein per fraction was determined by a Pierce™ BCA Protein Assay (Thermo Fisher Scientific).

3.2.2.2. IgA fractionation

One cartridge (8% Cartridge Kit, #42103, Abcam, Cambridge, UK) was set on the GELFrEE® 8100 fractionation system with HEPES 1X running buffer according to the manufacturer's instructions. Four

chambers were loaded with an IgA sample (25 μ L/chamber equivalent to 25 μ g of protein) in reducing and denaturing conditions (described for sample preparation in section 3.2.2.1). The electrophoresis was programmed to generate 12 fractions by controlling time and voltage and conducted according to the manufacturer's instructions. The fractions were stored at -20 °C. The 12 protein fractions were screened via SDS-PAGE 8-16% Tris-Glycine gel (Novex™ Tris-Glycine Protein Mini Gel 8-16%, XP08165BOX, Thermo Fisher Scientific, CA, USA). The first fraction showed a protein band of around 25k Da (IgA L-fraction). Fractions 7–9 were pooled since a band between 60 and 70 kDa was consistently observed (IgA H-fraction). The fractions 2–6 and 10–12 were discarded since no protein band was observed.

3.2.3. Proteolytic digestion

3.2.3.1. Blood plasma-derived samples

Tryptic digestion was performed using the FASP method [117], as described by Hoffmann *et al.* [118]. This digestion protocol was applied to the fractionated top 14-HAP depleted sample, the untreated blood plasma, and the top 14-HAP depleted sample using 60 μ g of protein. For all filtration steps, 10 kDa Nanosep® Omega filters (OD010C35, Pall®, Puerto Rico) were used. The reagents, 0.4 M DTT and 0.55 M IAA were dissolved immediately before use in 50 mM ABC_(aq) buffer pH 7.8 (ABC buffer) and, then, diluted 10 times in urea buffer (8 M Urea_(aq) and 100 mM Tris-HCl_(aq) pH 8.5) to obtain 40 mM DTT and 55 mM IAA. The filters were shaken and centrifuged at 14,000 x g for 10 min after each washing step. From each fraction derived from the GELFrEE® system, an amount of 60 μ g of protein was loaded onto the filter and washed twice with 200 μ L urea buffer. Then, a volume of 100 μ L 40 mM DTT was added and incubated for 20 min at 56 °C and 300 rpm on a ThermoMixer C (Eppendorf, Hamburg, Germany). After centrifugation at 14,000 x g for 10 min, 100 μ L of 55 mM IAA was added. The sample was incubated for 20 min in the dark, at room temperature and 300 rpm. Once reduced and carbamidomethylated, the protein sample was washed three times with 100 μ L urea buffer and three times with 100 μ L ABC buffer. Trypsin was added at a ratio of 1:60 (mg enzyme: mg protein sample) in a final volume of 100 μ L 50 mM ABC_(aq) + 5% (v/v) ACN + 1 mM CaCl_(aq). The sample was incubated overnight at 37 °C and 300 rpm (incubator Titramax 1000, Heidolph, Schwabach, Germany). The digest was recovered and collected by centrifugation (10,000 x g for 15 min). The membrane was washed once with 50 μ L of 50 mM ABC_(aq) + 5% (v/v) ACN and once with 50 μ L water. The flow-through was kept and combined with the digest. The digest was divided into aliquots with 20 μ g of peptides. The digests were dried using the rotational vacuum concentrator (0.01 mbar, ca. 3 h, 1°C, parameters applied for all drying steps) and stored at -20 °C.

3.2.3.2. IgA derived samples

The proteolytic digestion of the IgA samples and fractions was conducted as described in section 3.2.3.1. Only half of the volume of both IgA fractions (L- and H-) was required for proteolytic digestion (the other half was stored for future experiments). In addition, an amount of 100 µg of each IgA commercial sample (from supplier 1 and supplier 2, separately) was digested without prior protein fractionation. The trypsin per protein ratio and incubation conditions were described in section 3.2.3.1. The digests were dried in a rotational vacuum concentrator and stored at -20 °C.

3.2.4. Glycopeptide enrichment

3.2.4.1. Blood plasma-derived samples

Glycopeptide enrichment via cotton-HILIC-SPE was based on a protocol from Selman *et al.* [124]. Cotton tips were produced in-house by introducing a 3 mm long mercerized cotton thread at the bottom of 250 µL Rainin tips (Pipette Tips RT LTS 250µL SX 768A/8 Mettler Toledo, CA, USA). The washing solution (85% (v/v) ACN_(aq) 0.1% (v/v) TFA_(aq)) and first-elution solution (78% (v/v) ACN_(aq) 0.1% (v/v) TFA_(aq)) were prepared immediately before use. For each GELFrEE fraction, an amount of 20 µg of lyophilized tryptic digest was solubilized in 50 µL 85% (v/v) ACN_(aq). The thread-filled tips were cleaned thrice by pipetting 100 µL water and disposing of the volume. The tips were equilibrated by pipetting five times 100 µL washing-solution. The peptide sample was loaded into the cotton-tips by slowly pipetting the liquid 20 times up and down. The liquid from the loading step was discharged in a clean tube and kept (loading fraction). The tips were washed thrice by pipetting 100 µL washing solution into a clean tube (wash fraction). The elution 1 fraction was recovered by pipetting three times 100 µL first-elution solution into a clean tube. The elution 2 fraction was recovered by pipetting six times 100 µL H₂O into a clean tube. The four fractions were dried in a vacuum and stored at -20 °C. This Cotton-HILIC-SPE protocol was also applied to 20 µg of tryptic peptides from the untreated blood plasma and the top 14-HAP depleted sample. The HILIC fractions were solubilized in 20 µL 2% (v/v) ACN_(aq) 0.1% (v/v) TFA_(aq) on the day of injection. Four microliters of sample were injected per LC- MS/MS run.

3.2.4.2. IgA derived samples

The total amount of tryptic digest from the IgA L- and H-fractions was resuspended in 20 µL 85% (v/v) ACN_(aq). The tryptic peptides from the two unfractionated IgA samples (from suppliers 1 and 2) were resuspended in 100 µL 85% (v/v) ACN_(aq). For each unfractionated sample, the glycopeptide enrichment was conducted by four replicates using 20 µL of tryptic digest each time. The tryptic digest volume taken

(20 μL) was adjusted to 50 μL with 85% (v/v) $\text{ACN}_{(\text{aq})}$ for all cases, and the cotton-HILIC-SPE protocol described in section 3.2.4.1. was conducted. In the end, four HILIC-fractions were collected for each case: loading, wash, elution 1, and elution 2. Loading and wash fractions were pooled since both contain mostly unmodified peptides. All HILIC-fractions were dried, stored at $-20\text{ }^{\circ}\text{C}$, and resuspended in 10 μL water on the injection day. Three microliters of sample were injected per LC-MS/MS run (described in section 3.2.6.1.).

3.2.5. Glycoproteomic analysis of human blood plasma-derived samples

3.2.5.1. nano RP-LC-ESI-OT-OT-MS/MS data-dependent acquisition

Nano-reversed-phase liquid chromatography electrospray ionization orbitrap tandem mass spectrometry (nanoRP-LC-ESI-OT-OT-MS/MS) was performed using a Dionex UltiMate 3000 RSLCnano system (UHPLC, Thermo Fisher Scientific, Germering, Germany) coupled online to an Orbitrap Elite Hybrid Ion Trap-Orbitrap Mass Spectrometer (Thermo Fisher Scientific, Bremen, Germany). The UHPLC system was equipped with a C18 trap column (length 2 cm, particle size 5 μm , pore size 100 \AA , inner diameter 100 μm , Acclaim PepMap™ 100, #164199 Thermo Fisher Scientific, Lithuania) and a C18 separation column (length 25 cm, pore size 100 \AA , particle size 2 μm , inner diameter 75 μm , Acclaim PepMap™ RSLC nanoViper #164941, Thermo Fisher Scientific, Lithuania). For the separation gradient, the nano pump mobile phase A (2% (v/v) $\text{ACN}_{(\text{aq})}$, 0.1% (v/v) $\text{FA}_{(\text{aq})}$) and mobile phase B (80% (v/v) $\text{ACN}_{(\text{aq})}$, 10% (v/v) $\text{TFE}_{(\text{aq})}$, 0.1% (v/v) $\text{FA}_{(\text{aq})}$) were controlled at a flow rate of 300 nL/min at $40\text{ }^{\circ}\text{C}$. Then, 5 min after the sample injection, the loading pump mobile phase A (2% (v/v) $\text{ACN}_{(\text{aq})}$ 0.05% (v/v) $\text{TFA}_{(\text{aq})}$, flow rate 7 $\mu\text{L}/\text{min}$) switched the port valve to 1-2, connecting trap and separation column. The following separation gradient was set: 4% B (0–4 min); 4–35% (4–66 min); 35–90% (66–68 min); 90% (68–74 min); 90–4% in (74–76 min); 4% (76–100 min).

The orbitrap mass analyzer (OT-OT- MS/MS) acquired precursor and fragment ion scans. Each sample was measured twice to obtain data on the glycopeptides fragmented at two HCD regimes: HCD.low with fixed NCE of 20 and HCD.step with stepped NCE of 35 (width 15%, two steps), as described by Hoffmann *et al.* [118]. Both data-dependent scan methods were acquired in positive mode, selecting the top 5 peaks with 10 sec dynamic exclusion, scan range 350–2000 m/z , and isolation width 4 m/z . Orbitrap mass analyzer was used for precursor ion scan (MS^1) and fragment ion scan (MS^2), selecting 30,000 and 15,000 resolution for each scan, respectively.

3.2.5.2. Data analysis

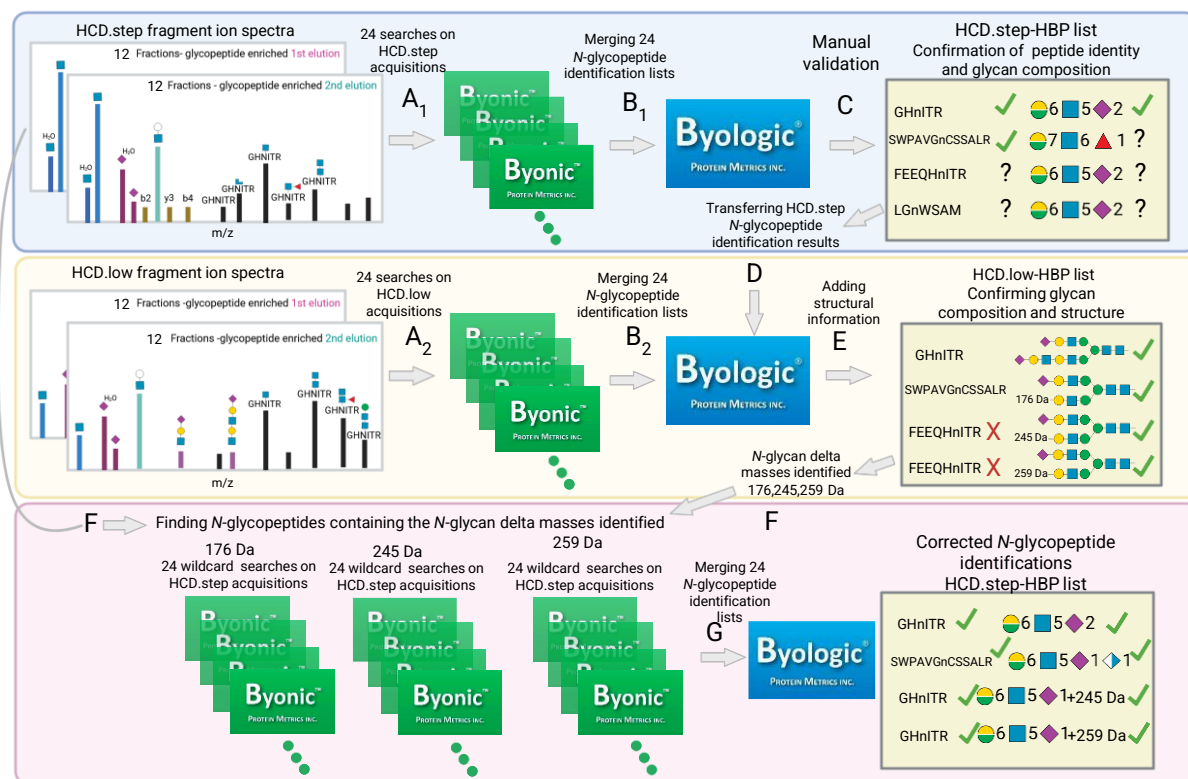


Figure 3.2. Human blood plasma N-glycoproteomic data analysis workflow. The intact N-glycopeptides, present in 24 glycopeptide-enriched fractions (12 x Elution-1 + 12 x Elution-2) were analyzed by LC-MS/MS using both fragmentation energies, HCD.step and HCD.low. (A) The spectra file acquired at each fragmentation regime was searched for N-glycopeptides using Byonic™ software. (B) All individual searches corresponding to the same fragmentation energy were combined using Byologic™. After this step, two large N-glycopeptide identification lists were produced: the HCD.step-HBP list and the HCD.low-HBP list. (C) The N-glycopeptides in the HCD.step-HBP list were manually validated to confirm the peptide and N-glycan composition assigned by the software. (D) The N-glycopeptide-identifications confirmed were imported into the HCD.low-HBP list to substitute corresponding precursor ions potentially incorrectly identified, using their mass and retention time. (E) Features annotated in the HCD.step-HBP list were transferred to the HCD.low-HBP list, and an additional revision for N-glycan structural evidence was conducted in the HCD.low-HBP list. (F) A second Byonic™ search focused on the new features identified was triggered. (G) Finally, a second HCD.step-HBP list containing corrected N-glycopeptide identifications was generated using Byologic™—image created with BioRender.com.

The HCD.step measurements of the HILIC-fractions (loading, wash, elution 1 and elution 2) derived from the three samples, untreated blood plasma, the top 14-HAP depleted sample, and the fractionated top 14-HAP depleted sample, were searched for proteins through the human UniprotKB/SwissProt database (v2021-11-30, 20,306 canonical sequences) using Proteome Discoverer™ (version 2.5.0.400, Thermo Fisher Scientific, Bremen, Germany). The parameters applied to the search engines Sequest HT and Mascot were set as follows: full specific tryptic digestion, two missed cleavages allowed, precursor and fragment mass tolerance: 10 ppm and 0.02 Da, respectively. Deamidation (N, Q), oxidation (M), and acetylation (protein N-Terminus) were established as dynamic modifications, and carbamidomethylation (C) as static

modification. The percolator algorithm was applied for peptide validation, and 1% FDR on peptide level was used.

The workflow of glycoproteomic data analysis is depicted in Figure 3.2. The Byonic™ search engine (v4.2.10 by Protein Metrics, Cupertino, CA, USA) was used to search for *N*-glycopeptides in the HCD.step and HCD.low measurements (Figure 3.2. steps A1 and A2). HCD.step generates abundant peptide fragment ions (a, b, y) and some fragments from the peptide-linked glycan moiety (Y ions). This fragmentation energy generates the following four characteristic *N*-glycopeptide Y ions: (I) $[M_{\text{peptide}+\text{H}} - \text{NH}_3]^+$, (II) $[M_{\text{peptide}+\text{H}}]^+$, (III) $[M_{\text{peptide}+\text{H}} + {}^{0.2}\text{X HexNAc}]^+$ and (IV) $[M_{\text{peptide}+\text{H}} + \text{HexNAc}]^+$ (where $\text{NH}_3=17.0265$ Da, ${}^{0.2}\text{X HexNAc}=83.0371$ Da; and $\text{HexNAc}=203.0794$ Da) [118, 141-143]. The HCD.low MS² spectra are rich in glycan fragment ions and Y ions with a more extended portion of the *N*-glycan attached to the peptide, which allows *N*-glycan structure interpretation. The following parameters were applied in Byonic™: specific tryptic digestion, maximum two missed cleavages, carbamidomethylation (C) as fixed peptide modification, and oxidation (M) as common-1 modification, deamidation (N), and pyro Glu/Gln (E) as rare-1 dynamic modifications. Precursor and fragment mass tolerance was set to 10 ppm and 20 ppm, respectively. The recalibration lock mass was 445.1201 m/z [230, 231]. Fragmentation-type was set to QTOF/HCD. The human canonical proteome UniProtKB/Swiss-Prot database (20,396 reviewed sequences downloaded in May 2021) was applied for protein identification [232]. A customized *N*-glycan database of 288 compositions was used for *N*-glycan identification [226]. During the Byonic™ search, no decoy search was performed, and there were no cuts on protein FDR. The *N*-glycoproteomic analysis of untreated blood plasma and top 14-HAP depleted sample (HCD.step acquisitions) was conducted using the same parameters.

The database with 288 entries of *N*-glycan compositions resulted from combining the “182 *N*-glycan Byonic™ database” with reported compositions, including multiple fucoses and compositions of phosphorylated or sulfated *N*-glycans deduced through a diagnostic search [226]. The diagnostic search focused on identifying fragment ions derived from phosphorylated or sulfated *N*-glycans and deducing the composition of such *N*-glycans. To perform the diagnostic search, two layout filters were created in Thermo Xcalibur Qual Browser (Version 2.2, Thermo Fisher Scientific, Bremen, Germany) software. The layouts were applied on each HCD.step file to filter MS² spectra containing oxonium marker ions of sulfated *N*-glycans $\text{HexNAc}_1\text{Sulfo}_1 [M+\text{H}]^+$ (Supplementary Figure A.1.1 Appendix) or $\text{HexNAc}_1\text{Hex}_1\text{Sulfo}_1 [M+\text{H}]^+$ or phosphorylated *N*-glycans $\text{Hex}_1\text{Phospho}_1 [M+\text{H}]^+$ or $\text{Hex}_2\text{Phospho}_1 [M+\text{H}]^+$ (Supplementary Figure A.1.2 Appendix). In the filtered MS² spectra, the presumable peptide mass was determined by detecting the four characteristic Y ions (I-IV). Then, the corresponding *N*-glycan mass was calculated by subtracting the peptide mass from the precursor ion mass. The sulfated and phosphorylated *N*-glycan compositions were

deduced using the GlycoMod tool from expasy.org [233]. The sulfated and phosphorylated *N*-glycans compositions were appended to the *N*-glycan database.

The search results derived from HCD.low and HCD.step acquisitions were imported into Byologic™ (v4.4-74-g75311a1df5 x64, Protein Metrics, Cupertino, CA, USA) in two sets (Figure 3.2. steps B1 and B2). The searches corresponding to HCD.step acquisitions generated one Byologic™ file (HCD.step-HBP), and the ones corresponding to HCD.low acquisitions generated a second Byologic™ file (HCD.low-HBP). For validation of *N*-glycopeptides identifications, a decision tree was established. Figure 3.2, in step C, refers to the manual validation step of the *N*-glycopeptides detected. As HCD.step typically generates the characteristic Y ions (I-IV), the validation was performed only on the list in the Byologic™ file HCD.step-HBP. The validated *N*-glycopeptide identifications in the HCD.step-HBP list were transferred to the HCD.low-HBP list (also a Byologic™ file) to substitute the corresponding correct precursor ion identifications in the HCD.low-HBP list (Figure 3.2. step D). This substitution was performed using the Peptide Manager Byonic™ function. The validated *N*-glycopeptide identifications in the HCD.step-HBP list were exported using the “Export in-silico to CSV” option. Then, the CSV file was imported into the HCD.low-HBP list through the Peptide Manager “Intersect CSV” function. The parameters set for intersecting the corresponding *N*-glycopeptides were precursor mass error (10 ppm) and retention time error (5 minutes). After substituting the correct gPSMs in the HCDlow-HBP list, the *N*-glycopeptide identifications were manually reviewed to annotate additional *N*-glycan structural information (Figure 3.2. step E).

Rare *N*-glycan compositions were identified during *N*-glycopeptide validation (Figure 3.2. step F). This relates to three *N*-glycan building block (supposedly monosaccharide masses) not included in the first *N*-glycan database. To find the gPSMs for such identifications, the Byonic™ wildcard search function was applied [156]. This function allows adding a delta mass within a user-specified range. The mass ranges searched here were: 176, 245, and 259 [± 1 Da]. These ranges were narrowed to the delta masses deduced during the validation (176.0314 Da, 245.0524 Da, and 259.0672 Da). One specific wildcard search was applied for each delta mass to all the HCD.step files [226]. To reduce the search space, MS/MS filtering function was used, allowing MS² spectra containing at least two of the following masses: HexNAc₁Hex₁ [M+H]⁺/366.1395, HexNAc₁ [M+H]⁺/204.0867, HexNAc₁ [M-H₂O+H]⁺/186.0761 or NeuAc₁ [M-H₂O+H]⁺/274.0921, mass tolerance 0.02 Da. In the wildcard search, the delta mass is assigned to the glycan modification, and not the peptide sequence by setting the parameter restrict to residues to “g” (where “g” means glycan). For both wildcard searches, the same parameters as the first Byonic™ search were applied: specific tryptic digestion, maximum 2 missed cleavages, carbamidomethylation (C) as fixed peptide modification, and oxidation (M) as common-1 modification, deamidation (N) and pyro Glu/Gln (E) as rare-1 dynamic modifications. Precursor and fragment mass tolerance was 10 ppm and 20 ppm, respectively.

The recalibration lock mass was 445.1201 m/z. Fragmentation-type set to “QTOF/HCD.” A shorter protein sequence list and *N*-glycan composition list, including only the elements in the “True”-validated *N*-glycopeptide results, were applied to speed up the search. The identifications resulting from all wildcard searches were integrated into the list HCD.step-HBP in Byologic™ (Figure 3.2., step G).

3.2.6. Glycoproteomic analysis of human IgA samples

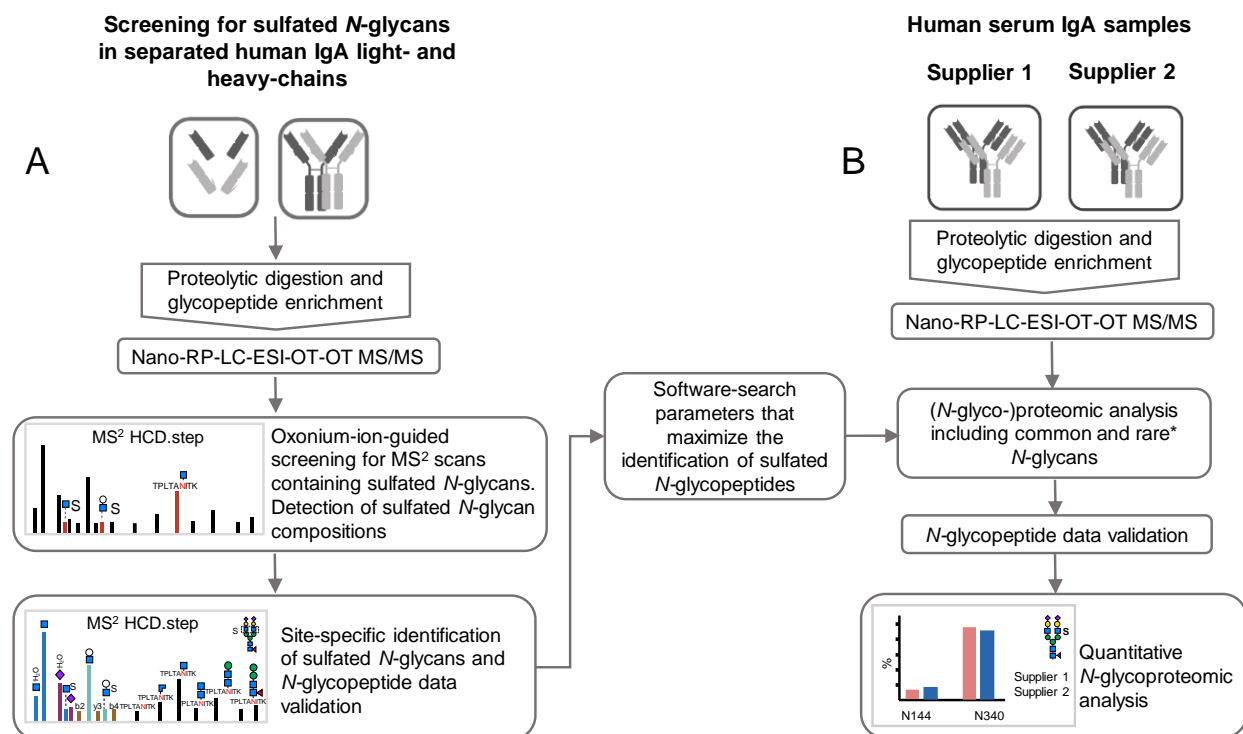


Figure 3.3. Experimental design for the site-specific identification of sulfated and other rare *N*-glycans in IgA. (A) First approach for the site-specific identification of sulfated *N*-glycans on fractionated human serum IgA. (B) Second approach via the analysis of unfractionated human serum IgA samples. Asterisk (*) indicates the application of additional strategies that can expand the *N*-glycoproteomic analysis of IgA. Nano-RP-LC-ESI-OT-OT-M/MS: Nano-reversed-phase liquid chromatography electrospray-ionization orbitrap tandem mass spectrometric measurement for precursor and fragment ions.

3.2.6.1. nano RP-LC-ESI-OT-OT-MS/MS data-dependent acquisition

NanoRP-LC-ESI-OT-OT-MS/MS was conducted using a Dionex UltiMate 3000 RSLCnano system coupled to an Orbitrap Eclipse Tribrid Mass Spectrometer (Thermo Fisher Scientific, Dreieich, Germany). A C18 trap column (length 2 cm, pore size 100 Å, particle size 5 µm, inner diameter 100 µm, Acclaim PepMap™ 100, #164199 Thermo Fisher Scientific, Lithuania), together with a C18 separation column (length 25 cm, pore size 100 Å, inner diameter 75 µm, particle size 2 µm, Acclaim PepMap™ RSLC nanoViper #164941, Thermo Fisher Scientific, Lithuania) were used for sample separation in the UHPLC system. For the run, 3 µL of the

sample was injected isocratically using the loading pump (100% mobile phase A: 0.1% (v/v) FA_(aq)) and a flow rate of 7 μ L/min. Five minutes after injection, the trap column was switched in line with the separation column. The nano pump was operated at a flow rate of 300 nL/min at 40 °C using mobile phase A (0.1% (v/v) FA_(aq)) and mobile phase B (80% (v/v) ACN_(aq) + 0.1% (v/v) FA_(aq)). The following separation gradient was established: 5% B (0–5 min); 5–31% (5–35 min); 31–44% (35–40 min); 44–95% (40–41 min); 95% (41–46 min); 95–5% (46–47 min); 5% (47–60 min). The ion source spray voltage was set to 2.55 kV, and the temperature of the ion transfer tube was set to 275 °C.

Mass spectrometric data acquisition of precursor and fragment ion spectra was conducted with an Orbitrap Eclipse Tribrid mass spectrometer. Resembling the MS acquisitions performed by Hoffmann *et al*, our MS method applied HCD fragmentation with stepped NCE [118]. The technique was called HCD.step (as in section 3.2.5.1.). Precursor ion scans were performed in positive ion mode using a scan range of m/z 350 to 1,500. The automatic gain control (AGC) target was set to “standard” with a resolution of 120,000 and a maximum injection time of 50 ms. The data-dependent acquisition mode was set to a 1.5 sec cycle time between precursor spectrum acquisitions. Dynamic exclusion was set as follows: exclusion duration to 10 sec, repeat count one, and mass tolerance of 10 ppm. Data-dependent fragment ion scans (MS²) were acquired performing HCD.step at a resolution of 60,000, scan range set to “normal” and isolation window set to m/z 1.6. In the case of detecting sulfated *N*-glycopeptides, an additional HCD fragmentation event was triggered by the detection of at least two out of the three selected oxonium marker ions (HexNAc₁Hex₁ [M+H]⁺/ 366.1395, HexNAc₁Sulfo₁ [M+H]⁺/ 284.0435, HexNAc₁Hex₁Sulfo₁ [M+H]⁺/ 446.0963) with mass accuracy of 10 ppm. HCD fragmentation with a fixed NCE of 20 “HCD.low” was applied at a resolution of 60,000 with a scan range set to “normal” and an isolation window of m/z 1.6. As a result of the different modified fragmentation regimes, HCD.low fragment ion scans complemented the information content of the HCD.step spectra derived from the sulfated *N*-glycopeptides [118].

3.2.6.2. Data analysis on the fractionated IgA sample

Oxonium-ion guided detection of MS² spectra containing sulfated HexNAc and identification of *N*-glycan compositions

The MS² spectra were screened for HexNAc-sulfated oxonium marker ions (HexNAc₁Sulfo₁ [M+H]⁺/ 284.0435 and Hex₁HexNAc₁Sulfo₁ [M+H]⁺/ 446.0963) using a scan filter in Thermo Xcalibur Qual Browser software (Version 2.2, Thermo Fisher Scientific, Bremen, Germany) as described in section 3.2.5.2 and display in Supplementary Figure A.1.1 (Appendix). The mass of the sulfated *N*-glycan moieties was estimated for each MS² scan by subtracting the observed peptide mass from the precursor mass. The putative peptide mass was determined using the conserved four characteristic *N*-glycopeptide Y ions

described in section 3.2.5.2. and also by Hoffmann *et al.* [118]. The putative peptide and *N*-glycan masses list for the selected MS² spectra is presented in Supplementary Table A.2.1 (Appendix). The compositions of the sulfated *N*-glycans were predicted by entering the estimated mass of the sulfated *N*-glycan moiety in the open-source GlycoMod tool from Expasy.org [233].

***N*-Glycoproteomic and proteomic data analysis of the IgA heavy chain fraction**

The first data analysis (Figure 3.3A) focused on identifying *N*-glycopeptides bearing sulfated *N*-glycans in the HILIC-fractions derived from the IgA H-fraction (fraction showing evidence for sulfated *N*-glycans). The data analysis included three steps: 1) identification of *N*-glycopeptides harboring sulfated *N*-glycans, 2) identification of *N*-glycopeptides harboring common *N*-glycans, and 3) identification of non-glycopeptides (proteomic search). To explain the HCD.step MS² spectra containing sulfated *N*-glycans, multiple glycoproteomic analyses were performed using the Byonic™ search engine (v4.4.2 by Protein Metrics, Cupertino, CA, USA). Supplementary Table A.2.2 (Appendix) describes the parameters for all searches applied in the three steps. Four protein sequences were primarily used on the *N*-glycoproteomic searches: IgA1 and IgA2 heavy constant chains, J-chain, and pIgR. Both the proteomic and *N*-glycoproteomic (“(*N*-glyco-)proteomic”) searches were conducted with the Byonic™ search engine. The results from the three searches and all raw files were imported in Byologic™ (v4.6-37-gda29558119 x64, Protein Metrics, Cupertino, CA, USA) to create one (*N*-glyco-)peptide identification list, which was manually validated. The software-search parameters related to the identification of sulfated *N*-glycans were applied on the subsequent analysis of IgA samples (Figure 3.3B).

3.2.6.3. Data analysis on two IgA commercial samples

The data analysis of the second approach displayed also in Figure 3.4A integrated not only information acquired from the first approach but also from additional strategies used for expanding the elucidation of IgA micro-heterogeneity and including other rare *N*-glycans (Figure 3.4.B).

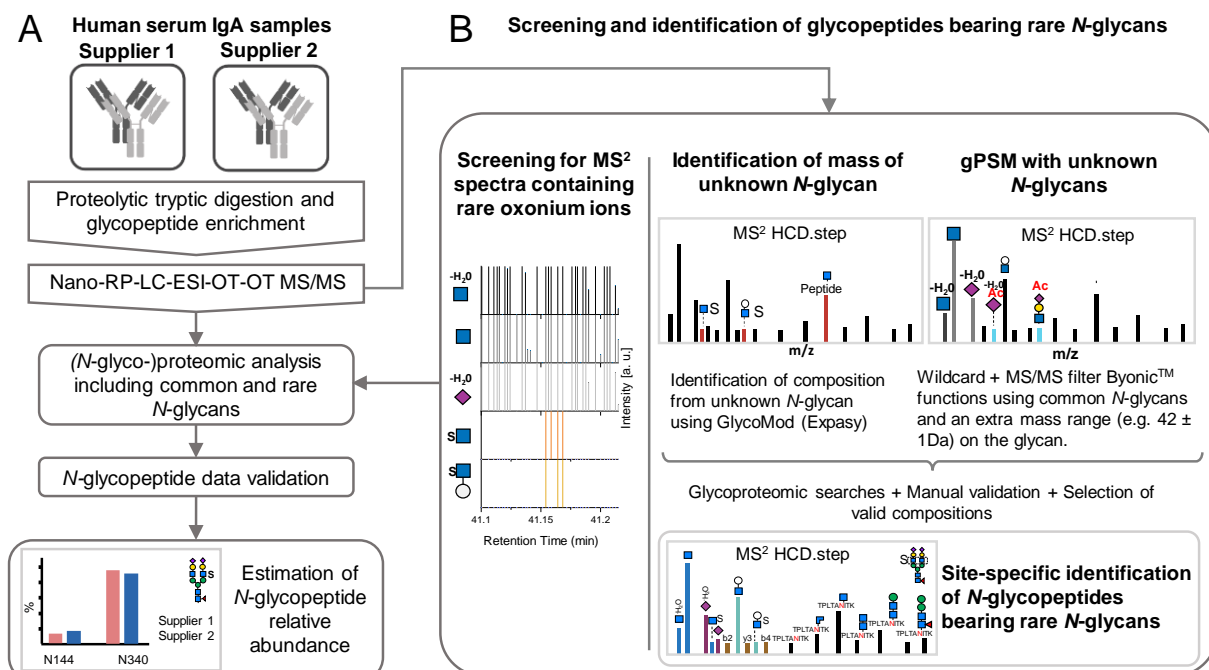


Figure 3.4. Expanding the N-glycoproteomic analysis of human serum IgA by applying multiple strategies. (A) Analysis of unfractionated human serum IgA of two commercial suppliers where common and rare N-glycans are integrated. (B) Strategies to determine the software search parameters for maximizing the identification of rare N-glycopeptides. Nano-RP-LC-ESI-OT-OT-MS/MS: nano-reversed phase-liquid chromatography with electrospray-ionization coupled orbitrap tandem mass spectrometric measurement for precursor and fragment ions. gPSM: glycopeptide spectra match.

Expanding the detection of rare N-glycans and identification of N-glycan compositions

As shown in Figure 3.4B, the two strategies relying on the characteristic oxonium marker ions were iteratively applied for the identification of new N-glycan compositions in acquired glycopeptide MS² spectra. The HexNAc-sulfated oxonium marker ions applied here were: HexNAc₁Sulfo₁ [M+H]⁺/ 284.0435 and Hex₁HexNAc₁Sulfo₁ [M+H]⁺/ 446.0963, while the NeuAc O-acetylated oxonium marker ions used here were: NeuAc₁Ac₁ [M-H₂O+H]⁺/ 316.1027 and HexNAc₁Hex₁NeuAc₁Ac₁ [M+H]⁺/ 699.2455). As mentioned before, the first strategy is based on estimating the mass of the unknown N-glycan moiety and predict the possible composition (based on oxonium ions observed) with the open source GlycoMod tool from expasy.org [233]. The mass of the unknown N-glycan moieties was estimated for each MS² scan by subtracting the observed putative peptide mass from the precursor mass. The second strategy implements both, glycan “wildcard” search and MS/MS filter Byonic™ functions [156]. This set up selectively conducts a glycoproteomic wildcard search on MS² spectra containing a defined set of oxonium marker ions. The wildcard search adds masses equivalent to a specific N-glycan modification (e.g., sulfate or acetyl group mass) to a set of common N-glycans and not to the peptide moiety.

Another goal of this work was screening for MS² scans containing other rare *N*-glycans present in the IgA samples, using the oxonium marker ions described in the next chapter (Table 4.2). As a result, *N*-glycopeptides presenting the oxonium ion HexNAc₁Hex₁HexA₁ [M+H]⁺ / 542.1716, where HexA is typically glucuronic acid in human *N*-glycans, were also detected. The correct *N*-glycan compositions were identified by means of the strategies presented above.

***N*-Glycoproteomic and proteomic data analysis**

The expanded data analysis (Figure 3.4) focused on the technical replicates from the two IgA samples (two suppliers) not fractionated by the GELFrEE® system. Both proteomic and *N*-glycoproteomic searches included the LC-MS/MS raw data from all HILIC-fractions (loading + wash, elution 1, and elution 2). One proteomic and eight glycoproteomic searches were conducted for each sample using the Byonic™ search engine. The proteomic search was set using the human canonical proteome UniprotKB (June 2022, 20386 canonical sequences). Then, the proteomic search generated a focused protein database, applied to the *N*-glycoproteomic searches. The parameters set for the (*N*-glyco-)proteomic searches per supplier sample are shown in Supplementary Table A.2.3 (Appendix). There, it is also described that four dedicated *N*-glycoproteomic searches supported the identification of *N*-glycopeptides: 1) *N*-glycan compositions with common *N*-glycans, 2) *N*-glycan compositions with HexNAc sulfation, 3) *N*-glycans compositions with *O*-acetylated sialic acid, and 4) *N*-glycan compositions with glucuronic acid. Four additional searches per sample focused on the identification of *N*-glycopeptides from truncated variants of the IgA1- and IgA2-tailpiece. Each dedicated *N*-glycoproteomic search was strictly conditionate by the presence of the oxonium marker ion characteristic of the *N*-glycan composition groups searched by using the MS/MS filter Byonic™ function. The results from the 18 searches were imported and combined in Byologic™ software, to create one (*N*-glyco-)peptide identification list. All *N*-glycopeptide identifications with a “Byonic™ MS² search score” above 100 were manually validated (as described in the subsequent section 4.2.2) and considered for further analyses. The non-glycopeptide identifications with a “Byonic™ MS² search score” above 100 were also trusted and considered for further analyses without manual validation. In this glycoproteomic analysis, gPSM (glycopeptide-spectrum match) presenting rare *N*-glycan compositions were accepted as “True” only if their corresponding oxonium marker ions were present in the MS² spectra.

To account for the instability of sulfated fragment ions and the effect of conditioning the identification of sulfated *N*-glycopeptides on the detection of sulfated oxonium ions, another glycoproteomic data analysis was performed on the MS/MS raw data of all HILIC-fractions from both samples. The analysis included only the searches for sulfated *N*-glycopeptides described in Supplementary Table A.2.3 (Appendix), but without setting HexNAc-sulfated oxonium ions (HexNAc₁Sulfo₁ [M+H]⁺ and Hex₁HexNAc₁Sulfo₁ [M+H]⁺) in the MS/MS filter Byonic™ function. The resulting *N*-glycopeptide identification lists were integrated in

Byologic™, manually validated. In this particular glycoproteomic analysis, gPSM lacking HexNAc-sulfated oxonium marker ions but correctly matching common oxonium ions, peptide b and y ions, and the conserved peptide fragmentation pattern were set as “True”, and considered for further comparative analyses.

Relative quantification of (*N*-glyco-)proteomic data

The relative abundance of IgA1 and IgA2 was calculated for all technical replicates of the IgA samples from both suppliers by including the LC-MS/MS raw files obtained from all HILIC-fractions. Label-free quantification was conducted within a single proteomic analysis using Proteome Discoverer™ (version 2.5.0.400, Thermo Fisher Scientific, Bremen, Germany). The software setup allowed the classification of the files by the supplier and technical replicates before the analysis. The search engines Sequest HT (Proteome Discoverer™ 2.5.0.400, Thermo Fisher Scientific, Bremen, Germany) and Mascot (version 2.6, Matrix Science, London, UK) were set using the human protein database SwissProt/UniprotKB (20,315 canonical sequences, v2022-06-14), full specific tryptic digestion, two missed cleavages allowed, precursor and fragment mass tolerance: 10 ppm and 0.02 Da respectively. The following dynamic modifications were set: deamidation (N, Q), oxidation (M), and acetyl (protein N-Terminus). Carbamidomethylation (C) was selected as a static modification. The algorithm percolator was applied for peptide validation, allowing 1% FDR on peptide level. Minora feature detector node, included in the processing workflow, enabled detecting and grouping peptide signals for the HILIC-fractions that belong to each technical replicate. The label-free quantification in the consensus workflow integrated the nodes feature mapped and precursor ions quantifier, considering “unique+razor” peptides with all other parameters set as default.

The proteomic data derived from Proteome Discoverer™ was processed with Microsoft Excel (2016). The relative abundance of IgA subclasses per replicate was calculated by normalizing the integrated peak area of each representative IgA1- or IgA2-peptide (“NFPPSQDASGDLYTTSSQLTLPATQCLAGK” or “NFPPSQDASGDLYTTSSQLTLPATQCPDGGK,” respectively) including variants due to missed cleavages and chemical modifications by the total integrated peak area of both peptides. The values of the technical replicates were averaged for each supplier sample.

Label-free quantification was also done on the validated (*N*-glyco-)peptide list from the second Byonic™ data analysis. The (*N*-glyco-)proteomic data combined in Byologic™ software was processed with Microsoft Excel (Professional Plus 2021) for relative quantification of *N*-glycan compositions per site (micro-heterogeneity). The (*N*-glyco-)peptides were classified by *N*-glycosylation site peptide sequence homology within the identified IgA sequences (P01876, P01877, and PODOX2 Uniprot June 2022), and then grouped by *N*-glycan compositions (neglecting differences caused by peptide modifications and missed

cleavages). The *N*-glycan compositions were quantified if at least three technical replicates showed values from valid *N*-glycopeptide identifications. The abundance of each *N*-glycan composition was normalized by the total area under the curve (AUC) values of the extracted ion chromatograms (XIC) of all *N*-glycan compositions found per the corresponding *N*-glycosylation site. The values of the technical replicate were averaged for each supplier sample.

4 In-depth analysis of the low-abundant *N*-glycoproteome of human blood plasma

Please note that parts of this section are taken from the original publication Zuniga-Banuelos *et al.* 2023 and 2025a [226, 228].

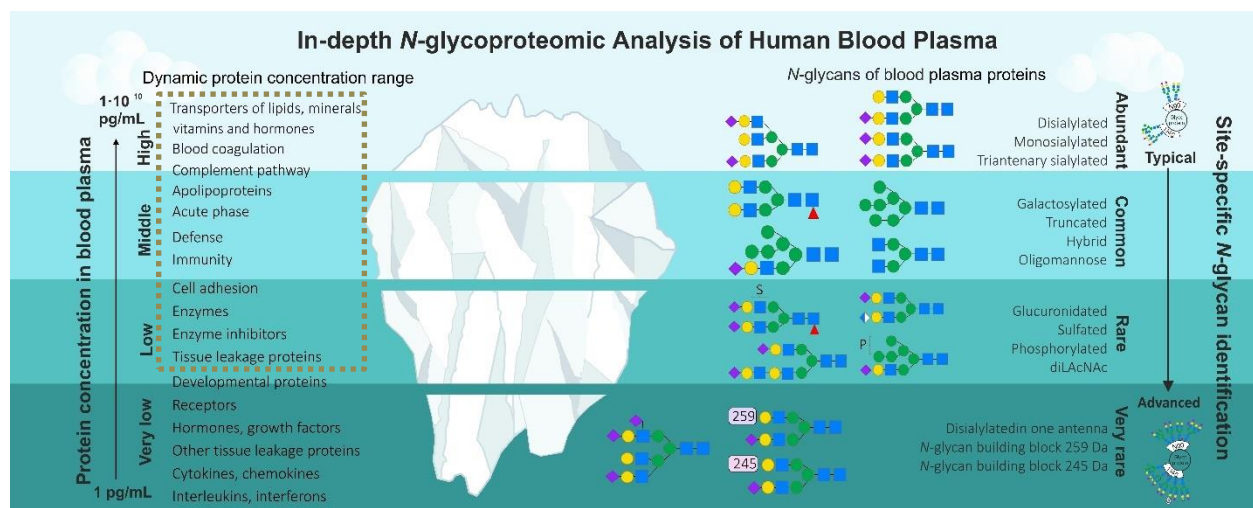


Figure 4.1. Exploring the low-abundant blood plasma glycoproteome. The *N*-glycan compositions detected by the developed in-depth *N*-glycoproteomic workflow are shown on the right. The dotted brown line indicates the range of proteins detected.

4.1. Introduction

Approximately 60% of the proteins reported as secreted to blood by the Human Protein Atlas [234, 235] are described as potentially glycosylated in the UniProtKB database [236, 237]. Growing evidence of the contribution of glycosylation to precision medicine enables the development of advanced glycan detection tools [238]. Further, technologies are evolving towards generating non-invasive early diagnostic platforms, which benefit from the elucidation of site-specific *N*-glycan features from the low-abundant glycoproteome. Nevertheless, advanced workflows are required first to extend the site-specific *N*-glycan identification to the lower abundance ranges. Clinical studies have revealed alterations in the *N*-glycosylation of human blood plasma proteins due to physiological and pathophysiological changes [6, 239, 240]. Glycoproteomic analysis via LC-MS/MS is the core technology for detecting such alterations of the *N*-glycosylation in blood plasma proteins in a site-specific manner [98]. In recent years, explorative large-scale glycopeptide-centered *N*-glycoproteomic studies on human plasma proteins, like IgG, enabled the identification of glycoforms that could serve as potential biomarkers or therapeutic targets [5, 6, 241, 242]. Several examples related to serum IgG were presented in section 2.3.3.1. Due to their organ specificity, promising blood plasma biomarker candidates are tissue-leakage proteins or signaling molecules [5].

Nevertheless, limitations related to the analysis of blood plasma and MS-based *N*-glycopeptide identification still impede the full diagnostic potential of *N*-glycoproteomic analyses. Firstly, the average concentration of tissue-leakage proteins in blood plasma is $1 \cdot 10^5$ – $1 \cdot 10^7$ times lower than the albumin concentration (35–50 mg/mL), the major blood plasma protein [24]. Due to the challenging concentration range, established workflows for blood plasma (glyco-)proteomics frequently include both a high-abundant protein depletion step and a fractionation step either on the peptide or protein level [30, 31]. A second constraint is that current MS-based workflows for intact *N*-glycopeptide identification give only a limited description of *N*-glycan structural information and an incomplete deduction of the input spectra. An example of these limitations' effect is glucuronidated *N*-glycans, detected in blood plasma via glycomics experiments but not yet reported through *N*-glycoproteomic analysis [29, 153]. This results in differences between glycoforms identified by glycomics and glycoproteomics analysis and points out the necessity to improve MS-based workflows for accurately describing the *N*-glycoproteome [129].

As explained in chapter 2.2.4., to conduct an optimal *N*-glycoproteomic search on a given LC-MS/MS dataset, the input search parameters (*N*-glycan list, protein list, modifications, etc.) must describe the characteristics of the different *N*-glycopeptides present in the sample. Furthermore, factors that neglect the existence of rare *N*-glycans (such as terminal glucuronic acid) and ignore post-glycosylational *N*-glycan

modifications such as sulfation, phosphorylation, and acetylation can lead to an inaccurate and inadequate description of the *N*-glycoproteome. To date, *N*-glycoproteomic data analyses tend to yield variable *N*-glycan and glycoprotein identifications when different *N*-glycopeptide search strategies and bioinformatics tools are applied and compared [148, 150]. The study from Kawahara *et al.* proposes that the main bottlenecks for obtaining precise results are the scoring algorithms and validation filters for gPSM [148]. Even though advanced FDR validation methods might improve the quality of gPSM, only a few software programs have demonstrated more efficient FDR validation methods after a controlled search [113, 243]. Thus, manual validation is still necessary for finding and reprocessing missed matches during an in-depth *N*-glycoproteomic data analysis, as was also concluded by Lee *et al.* [151].

Defining the *N*-glycan structure is another desired but underdeveloped aspect of the *N*-glycoproteomic analysis. The fragmentation of *N*-glycopeptides is favored by applying HCD-base methods, while for *O*-glycopeptides, EThcD-based methods are more useful [136]. In addition, diverse combinations of collisional energies used during fragmentation spectra acquisitions reflect specific advantages towards describing a glycopeptide (see chapter 2, section 2.2.4.2.) [131, 135]. A prominent setup is the HCD spectra acquired with a set of increasing NCE (e.g., 20, 35, and 50) [118]. As mentioned above HCD fragmentation with stepped energy is more efficient since it simultaneously generates information corresponding to both the peptide and *N*-glycan composition [118, 244].

In comparison, HCD spectra acquired at low NCE (e.g., fixed NCE 20) are rich in *N*-glycan structural information [118, 139]. As described in section 2.1.2., the *N*-glycan structure holds isomeric glycosidic bonds, compositions, anomeric configuration, and branches. By acquiring MS² spectra with HCD at both NCE, the wide variety of B and Y ions supports the differentiation of isomeric *N*-glycan branching structures at high confidence, as recently demonstrated by Hoffmann *et al.* and Shen *et al.* [118, 145]. Thus, it narrows, to some extent, the possible *N*-glycan structures associated with a determined *N*-glycan composition, for example, for distinguishing antenna or core fucose, bisecting HexNAc and diLacNAc [118]. Furthermore, detecting *N*-glycan structures holding post-glycosylational modifications, such as phosphorylation and sulfation, is critical for understanding glycoprotein-receptor interactions [67, 89, 245-249]. Detecting these rare *N*-glycan compositions and structures is an explorative task that can only be accomplished by manual annotation of *N*-glycan fragment ions, especially oxonium ion signals corresponding to rare or unknown *N*-glycan building blocks (e.g., glucuronic acid).

A preceding study in our group demonstrated the effectiveness of in-depth glycoproteomic analyses in maximizing the site-specific structural elucidation of *N*-glycans [118]. Furthermore, our prior established glycomic methods demonstrated reliability in the identification and characterization of rare *N*-glycans [45]. For example, the identification of a sulfated *N*-glycan [45], further described using a highly selective

sulfatase [44], prompts the search for rare *N*-glycans in the field of *N*-glycoproteomics. Thus, building on our previous work, it was integrated here a sample preparation workflow and further developed a data analysis workflow that allows in depth *N*-glycoproteomic analysis of human blood plasma. The present chapter reports the results of applying the workflow for an in-depth *N*-glycoproteomic analysis (depicted in section 3.2.) on a normal human blood plasma sample. As displayed in Figure 4.1, the preparative workflow not only enables the detection of glycoproteins within a concentration range of 7 orders of magnitude but also well characterizes the micro-heterogeneity of their *N*-glycans. The intact *N*-glycopeptide LC-MS/MS-based analysis, applied here, includes the acquisition of spectra using HCD.step and HCD.low fragmentation energies, which supports the differentiation of certain isomeric *N*-glycan branching structures. To conduct an in-depth exploration, manual validation was performed based on a newly developed decision tree. After the validation, a total of 1929 *N*-glycopeptides was identified, some of which bear *N*-glycan modifications such as sulfation and phosphorylation. Correcting invalid or partially valid *N*-glycopeptide identifications by *de novo* *N*-glycan sequencing disclosed rare *N*-glycan building blocks (like glucuronic acid). The insights and improvements gained will support further glycoproteomic software development and are crucial for future *N*-glycoproteomic studies aiming not only to explore biomarker candidates in human blood plasma, but also to evaluate therapeutic proteins and biological models.

4.2. Results

During in-depth *N*-glycoproteomic analysis of human blood plasma proteins, one encounters two leading causes of complexity. First, a few high-abundant proteins suppress the signal of hundreds of proteins present at lower abundance. Second, the heterogeneous and unpredictable nature of protein *N*-glycosylation requires sufficient evidence to describe the *N*-glycan structure and its position in a particular protein. In this study, a sample preparation and glycoproteomic data analysis workflow were established to reach the very low-abundant glycoproteome. Furthermore, it allowed the detection and site-specific description of rare *N*-glycan compositions and the identification of structural features linked to blood plasma glycoproteins with high confidence, as it will be substantiated in the following sections.

4.2.1. Exploring blood plasma (glyco-)proteins at low concentration range and *N*-glycan micro-heterogeneity

To evaluate the performance of the established preparative workflow for identifying middle- and low-abundant human blood plasma proteins, proteomic data analysis was conducted using Proteome Discoverer™ (see section 3.2.5.2.). The sample preparation workflow is presented in Figure 3.1, and its description is provided in Materials and Methods (chapter 3). The number of proteins identified was compared for each step of the sample preparation workflow (acceptance criterion: \geq two unique peptides per protein). The protein identifications were linked to the blood plasma concentrations reported in the Plasma Proteome Database (PPD) using the visProteomics R package [25, 250]. As presented in Figure 4.2A, the concentration of most of the proteins observed by analyzing the untreated blood plasma ranges from $1 \cdot 10^9$ down to $1 \cdot 10^6$ pg/mL. After depleting the top 14-HAP, the detected protein concentration range was extended now from $1 \cdot 10^9$ to $3 \cdot 10^5$ pg/mL. However, by integrating top 14-HAP depletion and protein molecular weight fractionation, the concentration range of the proteins identified could be further extended, and the detection limit further lowered, reaching now from $1 \cdot 10^9$ down to $6 \cdot 10^3$ pg/mL. Moreover, in a few cases detecting glycoproteins were detected at lower abundances, like the cysteine-rich secretory protein 3 (6.31 pg/mL) [25]. The list of proteins identified after each fractionation step is shown in Supplementary Tables A.1.1 to A.1.3 (Appendix).

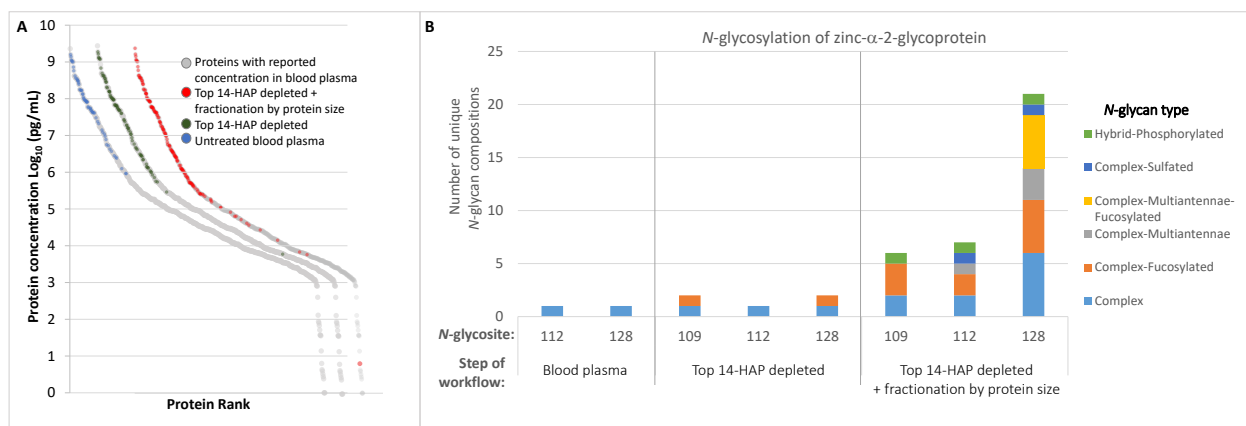


Figure 4.2. Gain achieved by applying the established sample preparation workflow. (A) Concentration range of the proteins identified after a proteomic analysis of the sample at each step of the preparative workflow (B) Measurement of micro-heterogeneity of glycosylation for zinc- α -2-glycoprotein (P25311) after a glycoproteomic analysis at each step of the preparative workflow.

Compared to the analysis of untreated blood plasma or the top 14-HAP depleted sample, the established sample preparation workflow also resulted in a significant improvement regarding the detection of more glycoproteins, glycopeptides and the measurement of micro-heterogeneity of glycosylation as displayed in Supplementary Figure A.1.3 (Appendix). Exemplarily, in Figure 4.2B, *N*-glycan compositions detected on each specific zinc- α -2 glycoprotein site are compared at each workflow step. For instance, as Table 4.1.

4. In-depth analysis of the low-abundant *N*-glycoproteome of human blood plasma

details, site N128, displayed ten times more *N*-glycan compositions after the top 14-HAP depletion than the direct analysis of blood plasma. When combining top 14-HAP depletion with protein molecular-weight fractionation, even 20 times more *N*-glycan compositions were achieved. Therefore, the preparative workflow effectively boosts the detection of a large variety of *N*-glycans at each site. A detailed list of the *N*-glycopeptides identified after each workflow step is shown in Supplementary Tables A.1.4 to A.1.6 (Appendix). An overview of the *N*-glycopeptide identification, after the in-depth *N*-glycoproteomic analysis, is displayed in the Supplementary Figure A.1.4 (Appendix). The plot shows that the most dominant *N*-glycan attached to blood plasma glycoproteins is a di-antennary complex-type *N*-glycan disialylated and non-fucosylated.

Table 4.1. Micro-heterogeneity of *N*-glycosylation observed in zinc- α -2-glycoprotein (P25311) after each step of the sample preparation workflow.

| Step of the workflow | Site | Type | <i>N</i> -glycan |
|--|------|-----------------------------------|-----------------------------------|
| Untreated blood plasma | 112 | Complex | HexNAc(4)Hex(5)NeuAc(2) |
| | 128 | Complex | HexNAc(4)Hex(5)NeuAc(2) |
| Top 14-HAP depleted | 109 | Complex-Fucosylated | HexNAc(4)Hex(5)Fuc(1)NeuAc(2) |
| | | Complex | HexNAc(4)Hex(5)NeuAc(2) |
| | 112 | Complex | HexNAc(4)Hex(5)NeuAc(2) |
| | 128 | Complex | HexNAc(4)Hex(5)NeuAc(2) |
| | | Complex-Fucosylated | HexNAc(4)Hex(5)Fuc(1)NeuAc(2) |
| | | Hybrid-Phosphorylated | HexNAc(3)Hex(7)NeuAc(1)Phospho(1) |
| Top 14-HAP plus protein molecular weight fractionation | 128 | Complex | HexNAc(4)Hex(5)NeuAc(2) |
| | | | HexNAc(4)Hex(5)NeuAc(2) |
| | | | HexNAc(4)Hex(5) |
| | | | HexNAc(4)Hex(5)NeuAc(1) |
| | | | HexNAc(4)Hex(4)NeuAc(1) |
| | | | HexNAc(4)Hex(6)NeuAc(1) |
| | | | HexNAc(4)Hex(5)NeuAc(2)Sulfo(1) |
| | | | HexNAc(5)Hex(6)NeuAc(3) |
| | | Complex-Multiantennae | HexNAc(5)Hex(6)NeuAc(2) |
| | | | HexNAc(5)Hex(6)NeuAc(1) |
| | | | HexNAc(5)Hex(6)Fuc(1)NeuAc(3) |
| | | Complex-Multiantennae-Fucosylated | HexNAc(5)Hex(5)Fuc(1)NeuAc(1) |
| | | | HexNAc(5)Hex(6)Fuc(2)NeuAc(2) |
| | | | HexNAc(5)Hex(7)Fuc(1)NeuAc(2) |
| | | | HexNAc(5)Hex(6)Fuc(1)NeuAc(2) |
| | | | HexNAc(4)Hex(5)Fuc(3)NeuAc(1) |
| | | Complex-Fucosylated | HexNAc(4)Hex(5)Fuc(1)NeuAc(1) |
| | | | HexNAc(4)Hex(4)Fuc(1)NeuAc(2) |
| | | | HexNAc(4)Hex(5)Fuc(1)NeuAc(2) |

4. In-depth analysis of the low-abundant *N*-glycoproteome of human blood plasma

| | | |
|-----|-----------------------|-----------------------------------|
| 112 | | HexNAc(4)Hex(5)Fuc(2)NeuAc(1) |
| | Hybrid-Phosphorylated | HexNAc(3)Hex(7)NeuAc(1)Phospho(1) |
| | Complex | HexNAc(4)Hex(5)NeuAc(2) |
| | | HexNAc(4)Hex(5)NeuAc(1) |
| | Complex-Fucosylated | HexNAc(4)Hex(5)Fuc(2)NeuAc(1) |
| | | HexNAc(4)Hex(5)Fuc(1)NeuAc(2) |
| | Complex-Sulfated | HexNAc(4)Hex(5)NeuAc(2)Sulfo(1) |
| | Complex-Multiantennae | HexNAc(5)Hex(6)NeuAc(3) |
| 109 | Hybrid-Phosphorylated | HexNAc(3)Hex(7)NeuAc(1)Phospho(1) |
| | Complex | HexNAc(4)Hex(5)NeuAc(2) |
| | | HexNAc(4)Hex(5)NeuAc(1) |
| | Complex-Fucosylated | HexNAc(4)Hex(5)Fuc(3)NeuAc(1) |
| | | HexNAc(4)Hex(5)Fuc(2)NeuAc(1) |
| | | HexNAc(4)Hex(5)Fuc(1)NeuAc(2) |

Overall, by comparing the (glyco-)proteomic results after each step of the preparative workflow, it is evident that the preparative workflow applied is expanding the range for identification of blood plasma (glyco-)proteins from middle to very low-concentration range and enabling a deeper description of the micro-heterogeneity of *N*-glycosylation.

4.2.2. Development of a data analysis workflow for the identification of intact *N*-glycopeptides

The dedicated data analysis workflow was developed and applied to validate the data generated by in-depth *N*-glycoproteomic LC-MS/MS analysis (see Figure 3.2. and section 3.2.5.2.). The workflow relies on revising two *N*-glycopeptide lists grouped by the fragmentation energy applied: HCD.step and HCD.low. The first part of the workflow uses HCD.step spectra to confirm the correctness of the peptide sequence and the *N*-glycan composition within the gPSM suggested by Byonic™. HCD.low fragment ion spectra are used to verify the proposed *N*-glycan composition in the second part of the data analysis workflow. Information about the *N*-glycan structure is manually added to clarify, for example, if fucose is linked to the core or the antenna of an *N*-glycan. After acknowledging errors, such as missed *N*-glycan compositions, the third part of the workflow focuses on correcting gPSM that were classified as uncertain identifications by *N*-glycan *de novo* sequencing or performing additional glycoproteomic searches.

The first part of the data analysis workflow was conducted according to the proposed decision tree shown in Figure 4.3A. With this decision strategy, a total of 7,867 gPSM were classified into three main categories: “True,” “Uncertain,” and “False.” A “True” gPSM has evidence on three aspects: 1) it is an *N*-glycopeptide,

2) the mass of the peptide moiety is consistent with the Y ion $[M_{\text{peptide}}+H]^+$, and 3) the oxonium ions observed are coherent with the *N*-glycan composition suggested. An “Uncertain” gPSM meets the first requirement but fails in (2) or (3) or both. A “False” gPSM does not fulfill the first requirement since it corresponds to a non-glycosylated peptide or an *O*-glycopeptide. The latter is recognized by the ratio between the intensity of HexNAc₁ $[M-H_2O+H]^+$ and HexNAc₁ $[M+H]^+$ oxonium ions [118]. In an *N*-glycopeptide-derived MS² spectrum (HCD.step), the intensity of HexNAc₁ $[M+H]^+/204.0867$ will be between three to ten times the intensity of HexNAc₁ $[M-H_2O+H]^+/186.0761$ [118]. As described in Figure 4.3A, each main category acquires a more specific classification according to the evidence observed.

Examples with a detailed description of each subcategory are displayed in Supplementary Figures A.1.5 to A.1.15 (Appendix). For example, the MS² spectra from a gPSM classified as “True-Evidence” contain at least three characteristic Y ions and oxonium ions consistent with the *N*-glycan composition presented in the gPSM; yet, the MS² spectra might lack peptide b and y ions. The category “Uncertain-change glycopeptide” is also an *N*-glycopeptide identification; however, the mass of the characteristic Y ions disagrees with the one expected from the gPSM suggested by the software.

4. In-depth analysis of the low-abundant *N*-glycoproteome of human blood plasma

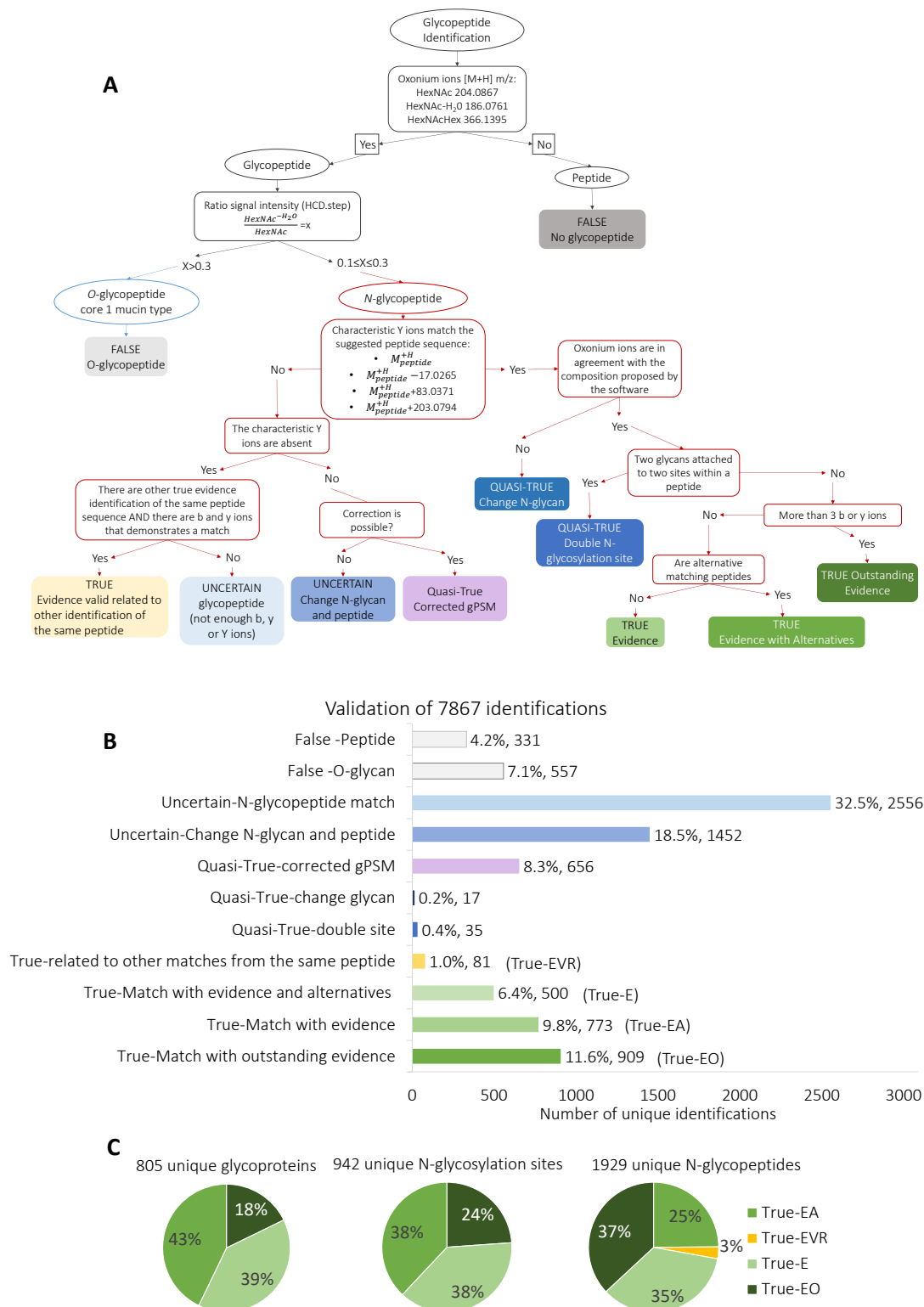


Figure 4.3. Validation of *N*-glycopeptide identifications. (A) Newly developed decision tree depicting the process for validating the gPSMs identified in the HCD.step-HBP list. (B) Number of gPSM assigned to each validation category. (C) The proportion of unique glycoproteins, *N*-glycosylation sites, and *N*-glycopeptides obtained from the true gPSM. T-EO: True Outstanding evidence. T-E: True-Evidence. T-EA: True-Evidence and Alternatives. T-EVR-True Evidence Valid Related to other identification of the same peptide.

4.2.3. Validated *N*-glycopeptide identifications in human blood plasma

Figure 4.3B shows the results of the manual validation of the HCD.step-HBP list. The “False” cases (non-*N*-glycopeptides) represent only 11% of the dataset. The results from the validation of HCD.step-HBP are listed in Supplementary Table A.1.6 (Appendix). From the total of 7,867 gPSMs, 2263 were correct concerning peptide and *N*-glycan composition (27.9% of the total gPSMs). The entire “True” gPSMs are spread across four subcategories: 909 “True matches with outstanding evidence” (T-EO), 773 “True matches with evidence” (T-E), 500 “True matches with evidence and alternative identifications” (T-EA) and 81 “True matches with evidence related to other matches from the same peptide sequence” (T-EVR). One-third of the total identifications correspond to “Uncertain-no-evidence” identifications, in which gPSM feature MS² spectra with a poor number of fragment ions. Another 17% of all gPSMs are incorrect matches (category “Uncertain-change *N*-glycan and peptide”), which MS² spectra might require different glycoproteomic searches. Some “Uncertain-*N*-glycan and peptide” were corrected by comparing their MS² spectra with the ones from “True” gPSM with similar precursor ion mass or retention time. Thus, 656 manually corrected gPSM resulted in the “Quasi-true corrected gPSM” subcategory. Other “Quasi-true” categories, “Change glycan” and “Double site,” represent only 0.6% of the total gPSMs and are cases in which the peptide moiety is correct, but the *N*-glycan cannot be confirmed. Figure 4.3C shows 942 different *N*-glycosites belonging to 805 human glycoproteins were found. Focusing on the different glycoforms per glycosylation site, the total number of gPSM is condensed to 1,929 by disregarding the peptide modifications or missed cleavages (Figure 4.3C).

4.2.4. Annotating structural features of bisecting, fucosylated, and LacNAc extended *N*-glycans using HCD.low spectra

Only the 2263 “True” gPSM were applied in the second part of the workflow to annotate *N*-glycan structural information in the HCD.low-HBP list (Supplementary Table A1.7 Appendix). *N*-glycan structural evidence was screened in the HCD.low spectra corresponding to the validated gPSM using the *N*-glycan marker ions listed in Table 4.2 [4, 118, 139, 251, 252]. This table considers Y ions in different charge states ($z=1-3^+$). This is the second part of the data analysis workflow depicted in Figure 3.2., which enables the differentiation between antenna and core fucose, multi-antennary *N*-glycans and repeated LacNAc units, or antenna HexNAc and bisecting HexNAc [118]. For instance, detecting the Y-ion peptide+HexNAc₃Hex₁ ion suggests a bisecting *N*-glycan. In total, 16 bisected *N*-glycopeptides were identified in 11 glycoproteins. Most of these bisected *N*-glycans are linked to the proteins IgA2 (heavy chain), IgG1 (heavy chain), IgG2 (heavy chain), and synaptojanin-1. In the case of a LacNAc repeat unit, the observation of the oxonium-ion HexNAc₂Hex₂NeuAc₁ (B-ion [M+H]⁺/1022.3671) was detected and annotated in the HCD.low-HBP list.

It has been reported that HexNAc₂Hex₂NeuAc₁ oxonium ion is not present in MS² spectra derived from complex-type *N*-glycans fragmented at low collision energy, hence providing high specificity towards identifying diLacNAc structures [139]. Of note, the oxonium ion HexNAc₂Hex₂ (B-ion [M+H]⁺/731.2717) was not applied because this fragment ion is also generated by complex-type and bisected *N*-glycans. Therefore, the identification of non-sialylated LacNAc repeat units was not included. It was found that apolipoprotein D harbors three different *N*-glycans with a sialylated LacNAc repeat unit at glycosylation site N65, the protein with the highest frequency of this type of *N*-glycan. Three additional *N*-glycopeptides belonging to kininogen-1, β -2-glycoprotein 1, and kallistatin were also found to have LacNAc repeat units.

Table 4.2. *N*-Glycan-derived marker ions. B and Y ions applied in this study

| <i>N</i> -glycan building block | Elemental composition (dehydrated) | Theoretical mass (Da) | Symbol |
|--|--|--|--------|
| ^{0.2} X HexNAc | C ₄ H ₅ NO | 83.0371 | |
| HexNAc | C ₈ H ₁₃ O ₅ N | 203.0794 | |
| Hex | C ₆ H ₁₀ O ₅ | 162.0528 | |
| NeuAc | C ₁₁ H ₁₇ O ₈ N | 291.0954 | |
| Fuc | C ₆ H ₁₀ O ₄ | 146.0579 | |
| GlcA | C ₆ H ₈ O ₆ | 176.0321 | |
| Phosphorylation | H ₁₀ O ₃ P | 79.9663 | P |
| Sulfation | O ₃ S | 79.9568 | S |
| Structural feature | Marker ion | Theoretical mass [+H ⁺] | Symbol |
| Bisecting-HexNAc | Peptide-HexNAc ₃ Hex ₁ | M _{peptide} ^{+H} + 771.2909 | |
| | Peptide-HexNAc ₃ Hex ₂ | M _{peptide} ^{+H} + 933.3438 | |
| Bisecting-HexNAc (core fucose) | Peptide-HexNAc ₃ Hex ₁ Fuc ₁ | M _{peptide} ^{+H} + 917.3488 | |
| | Peptide-HexNAc ₃ Hex ₂ Fuc ₁ | M _{peptide} ^{+H} + 1079.4017 | |
| Antenna fucose | HexNAc ₁ Hex ₁ Fuc ₁ | 512.1974 | |
| Sialylated antenna fucose | HexNAc ₁ Hex ₁ Fuc ₁ NeuAc ₁ | 803.2928 | |
| Core fucose | Peptide-HexNAc ₁ Fuc ₁ | M _{peptide} ^{+H} + 349.1373 | |
| | Peptide-HexNAc ₂ Fuc ₁ | M _{peptide} ^{+H} + 552.2167 | |
| | Peptide-HexNAc ₂ Hex ₁ Fuc ₁ | M _{peptide} ^{+H} + 714.2695 | |
| | Peptide-HexNAc ₂ Hex ₂ Fuc ₁ | M _{peptide} ^{+H} + 876.3223 | |
| Sialylated LacNAc repeated unit | HexNAc ₂ Hex ₂ NeuAc ₁ | 1022.3671 | |
| Two sialic acids per antenna | HexNAc ₁ Hex ₁ NeuAc ₂ | 948.3303 | |
| | HexNAc ₁ NeuAc ₁ | 495.1821 | |
| Glucuronidated antenna | HexNAc ₁ Hex ₁ HexA ₁ | 542.1716 | |
| <i>N</i> -glycan building block 245-Da antenna | HexNAc ₁ Hex ₁ 245 ₁ | 611.1899 (observed) | |
| <i>N</i> -glycan building block 259-Da antenna | HexNAc ₁ Hex ₁ 259 ₁ | 625.2049 (observed) | |
| Hexose phosphorylation | Hex ₁ Phospho ₁ | 243.0264 | |
| | Hex ₂ Phospho ₁ | 405.0793 | |
| | Hex ₃ Phospho ₁ | 567.1320 | |
| HexNAc sulfation | HexNAc ₁ Sulfo ₁ | 284.0434 | |
| | HexNAc ₁ Sulfo ₁ Hex ₁ | 446.0963 | |

*To facilitate the annotation glucuronic acid is represented as the type of hexuronic acid assumed to exist in the *N*-glycan.

In the HCD.step-HBP list, 679 unique fucosylated *N*-glycopeptides were further reviewed for deducing the position of fucose in the respective *N*-glycans (Supplementary Table A.1.6 Appendix). The presence of the following marker ions was registered in the HCD.step and HCD.low-HBP lists: for core-fucose, peptide+HexNAc₁Fuc₁[M+H]⁺ (Y-ion); and for antenna fucose, HexNAc₁Hex₁Fuc₁[M+H]⁺ (oxonium ion, B-ion). All in all, a total of 350 unique fucosylated *N*-glycopeptides having MS² spectra with either a core or an antenna fucose marker ion (Supplementary Table A.1.8 Appendix). Ambiguous cases occurred when only one fucose was suggested in the *N*-glycopeptide identification and both classes of marker ions, antenna- and core-fucose, were observed. Since fucose transfer (fucose rearrangement) in the gas phase can occur from core to antenna, generating “false” antenna fucose marker ions [253], the *N*-glycopeptides, including one fucose but presenting both antenna- and core-fucose ion were set as “core- or antenna-fucose”. In the case of *N*-glycan compositions with more than one fucose showing both antenna- and core-fucose ions, both core and antenna fucosylation were granted. Only antenna fucosylation was accepted for *N*-glycopeptides where core-fucose ions were not observed, but antenna-fucose ion was detected. As a result, 86 ambiguous core- or antenna-monofucosylated *N*-glycopeptides, 156 core-monofucosylated *N*-glycopeptides, 47 core- and antenna-fucosylated *N*-glycopeptides, and 61 *N*-glycopeptides presenting only one or more antenna-fucose were found. The proteins with the highest frequency on antenna-fucosylation are zinc- α -2-glycoprotein, α -2-HS-glycoprotein, β -2-glycoprotein 1, clusterin, and ceruloplasmin.

In Figure 4.4, all gPSM, including fucose(s) in the *N*-glycan composition, were plotted to investigate the frequency of fucosylated *N*-glycopeptides with and without fucose marker ion evidence. Many difucosylated gPSMs, which lack any fucose ion, were observed to present complex-type di-antennary monosialylated or tri-antennary disialylated *N*-glycans.

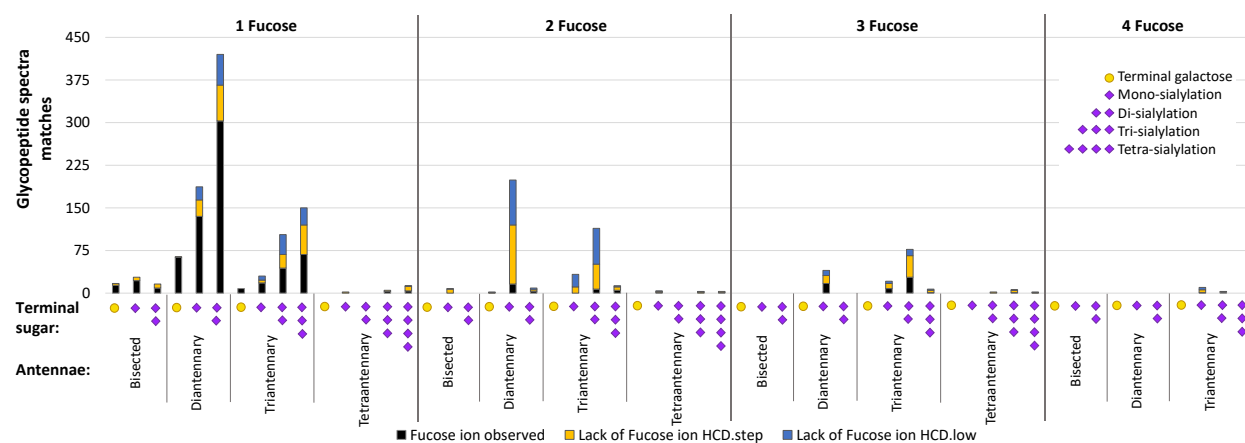


Figure 4.4. Evaluation of the evidence present in fucose-containing gPSMs. Distribution of all gPSMs that include fucose identified by Byonic™ software in both HCD.step and HCD.low-HBP lists (with and without fucose ion evidence observed).

A prominent example of difucosylated gPSMs is shown in Figure 4.5A, where no antenna-fucose oxonium ions are observed in the annotation done by the software. By analyzing the isotopic pattern of the triply charged precursor of this example (Figure 4.5B), it is observed an initial isotopic peak with an m/z difference of 0.3331 concerning the monoisotopic peak assigned by the software. This presumes that the mass of the precursor ion deduced by the software is one dalton larger than the real mass of such a precursor ion ($\Delta m/z=0.3331$, $z=3$, $\Delta m=0.9993$ Da). Then, the mass difference between assigning 2xFucoses instead of 1xNeuAc is one dalton (2xFucose=292.1158 Da, 1xNeuAc=291.0954 Da, $\Delta +1.0204$ Da). Hence, it was hypothesized that the software misrepresented the MS^1 isotopic pattern since it must start from a previous isotopic peak (Figure 4.5D). This would explain the mass difference of one dalton, which leads to assigning incorrectly two fucoses instead of one neuraminic acid. To confirm this, Figure 4.5C shows the manual *de novo* sequencing of the *N*-glycan composition reflected in the HCD.low spectrum from this precursor ion. As fucose ion evidence was not observed, neutral losses corresponding to fucose were searched. The result shows only a delta mass corresponding to a doubly charged NeuAc and a neutral loss from a second NeuAc at the end of the m/z range. This demonstrates that this *N*-glycan contains two NeuAc instead of one NeuAc and two fucoses. Other examples in which the potential real monoisotopic peak of the precursor ion is ignored in gPSM with multiple fucoses are shown in Supplementary Figures A.1.16 and A.1.17 (Appendix).

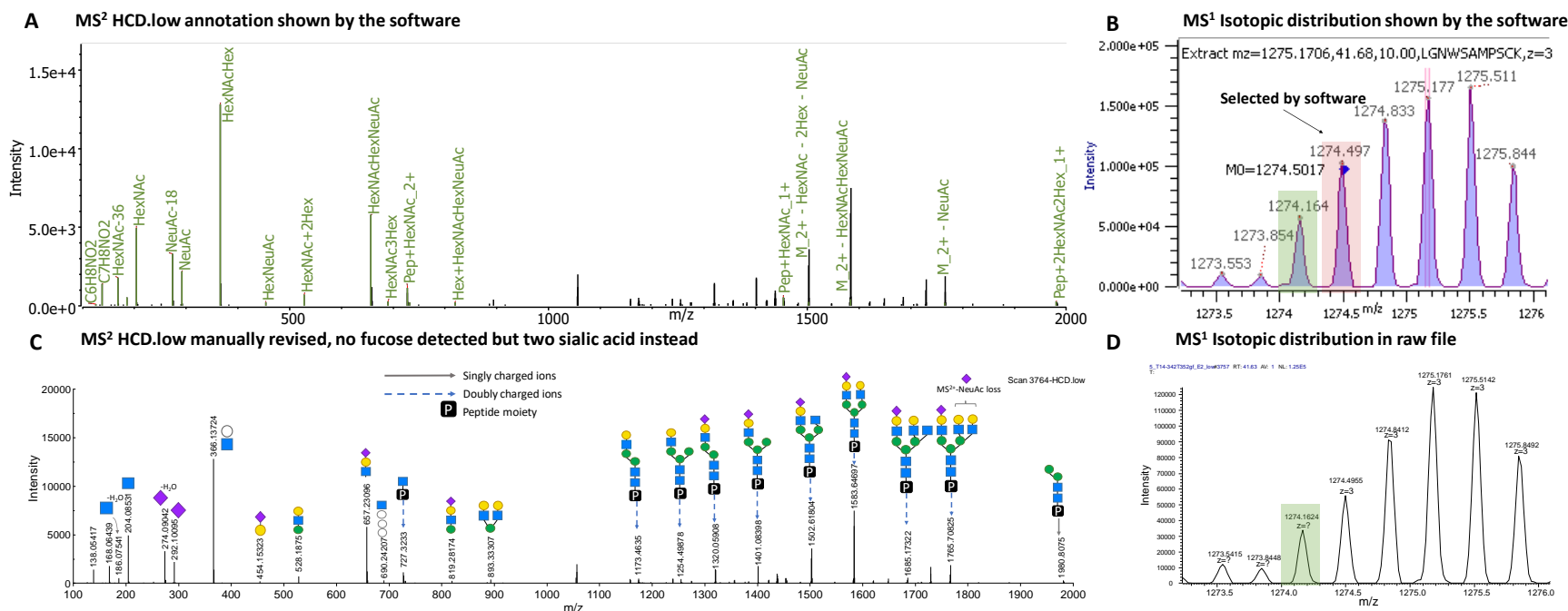


Figure 4.5. Manual revision of an incorrect identification for β -2-glycoprotein 1 (peptide K.LGNWSAMPSCK.A) containing a difucosylated complex-type N-glycan. (A) MS² HCD.low annotation proposed by ByonicTM software assigning an incorrect N-glycan HexNAc(5)Hex(6)Fuc(2)NeuAc(1) N-glycan mass: 2570.9250 Da. (B) MS¹ Isotopic distribution with monoisotopic peak selected by ByonicTM software (red). Calculated precursor ion m/z 1274.5017, $z=3$, observed mass 3820.4789 Da. Correct precursor ion highlighted in green. (C) MS² HCD.low manual annotation assigning the correct N-glycan HexNAc(5)Hex(6)NeuAc(2) N-glycan mass: 2569.9046 Da. (D) MS¹ Isotopic distribution observed in the raw file. Correct precursor ion highlighted in green. Precursor ion suggested m/z 1274.1624, $z=3$, calculated mass 3819.4872 Da.

4.2.5. Detection of sulfated and phosphorylated *N*-glycopeptides in common human blood plasma glycoproteins

N-Glycopeptides with sulfated and phosphorylated *N*-glycans are difficult to detect and locate due to their low abundance and unstable behavior during MS analysis [29, 185]. Therefore, a dedicated diagnostic search using Thermo Xcalibur Qual Browser was performed, as described in section 3.2.5.2. from chapter 3. Two layout-filters, shown in Supplementary Figures A.1.1 and A.1.2. (Appendix), were applied to detect *N*-glycopeptides with phosphorylated and sulfated *N*-glycan compositions in the human blood plasma. As a result, a collection of all possible phosphorylated and sulfated *N*-glycan compositions was added to the Byonic™ *N*-glycan composition database. This database was then used for all *N*-glycopeptide searches. Additionally, during manual validation, both the HCD.step- and HCD.low-HBP lists were revised concerning marker ions related to sulfation or phosphorylation. The example of Figure 4.6 shows that the HCD.low MS² spectrum of a sulfated *N*-glycopeptide identification contained the HexNAc₁Hex₁Sulfo₁ [M+H]⁺ oxonium marker ion, while the HCD.step MS² spectrum of the same *N*-glycopeptide displays the HexNAc₁Sulfo₁ [M+H]⁺ oxonium marker ion. The annotation displayed by Byonic™ software is shown in Supplementary Figure A.1.18 (Appendix). Interestingly, in none of the acquired MS² spectra, the Hex₁Sulfo₁ [M+H]⁺ oxonium ion was detected. With regard to phosphorylated *N*-glycopeptides (Figure 4.7), the HCD.low MS² spectrum mainly shows the Hex₁Phospho₁ [M+H]⁺ and Hex₂Phospho₁ [M+H]⁺ oxonium marker ions, while the HCD.step MS² spectra dominantly produce the Hex₁Phospho₁ [M+H]⁺ oxonium marker ion. The software annotation is shown in the Supplementary Figure A.1.19 (Appendix). Among the *N*-glycopeptide identifications from the HCD.step-HBP list, 65 contain sulfated *N*-glycans. Still, only 10 of them contain sulfation marker ions (Supplementary Table A.1.9 Appendix). In the HCD.step-HBP list, 15 *N*-glycopeptides presented phosphorylated *N*-glycans, and in contrast to the sulfated *N*-glycopeptides, all the MS² spectra contain marker ion(s) for phosphorylated hexose (Supplementary Table A.1.10 Appendix).

Figures 4.6 and 4.7 show that via *de novo* sequencing of the HCD.low MS² spectra, sulfation was detected on the antenna-HexNAc, and phosphorylation after the fourth mannose of the hybrid-type *N*-glycan. In the lists of validated gPSMs containing sulfation or phosphorylation (Supplementary Tables A.1.9 and A.1.10 Appendix), it is observed that the sulfated *N*-glycan compositions are related to complex-type *N*-glycans, while phosphorylated *N*-glycans are related to hybrid-type or oligomannose-type *N*-glycan. The *N*-glycosylation micro-heterogeneity of some proteins, like ceruloplasmin, hemopexin, heparin cofactor 2, and zinc α-2-glycoprotein, reflected several *N*-glycosylation sites harboring a sulfated *N*-glycan (Supplementary Table A.1.9 Appendix). The proteins ceruloplasmin, hemopexin and zinc α-2-glycoprotein also contained a higher frequency of phosphorylated *N*-glycans (Supplementary Table A.1.10 Appendix).

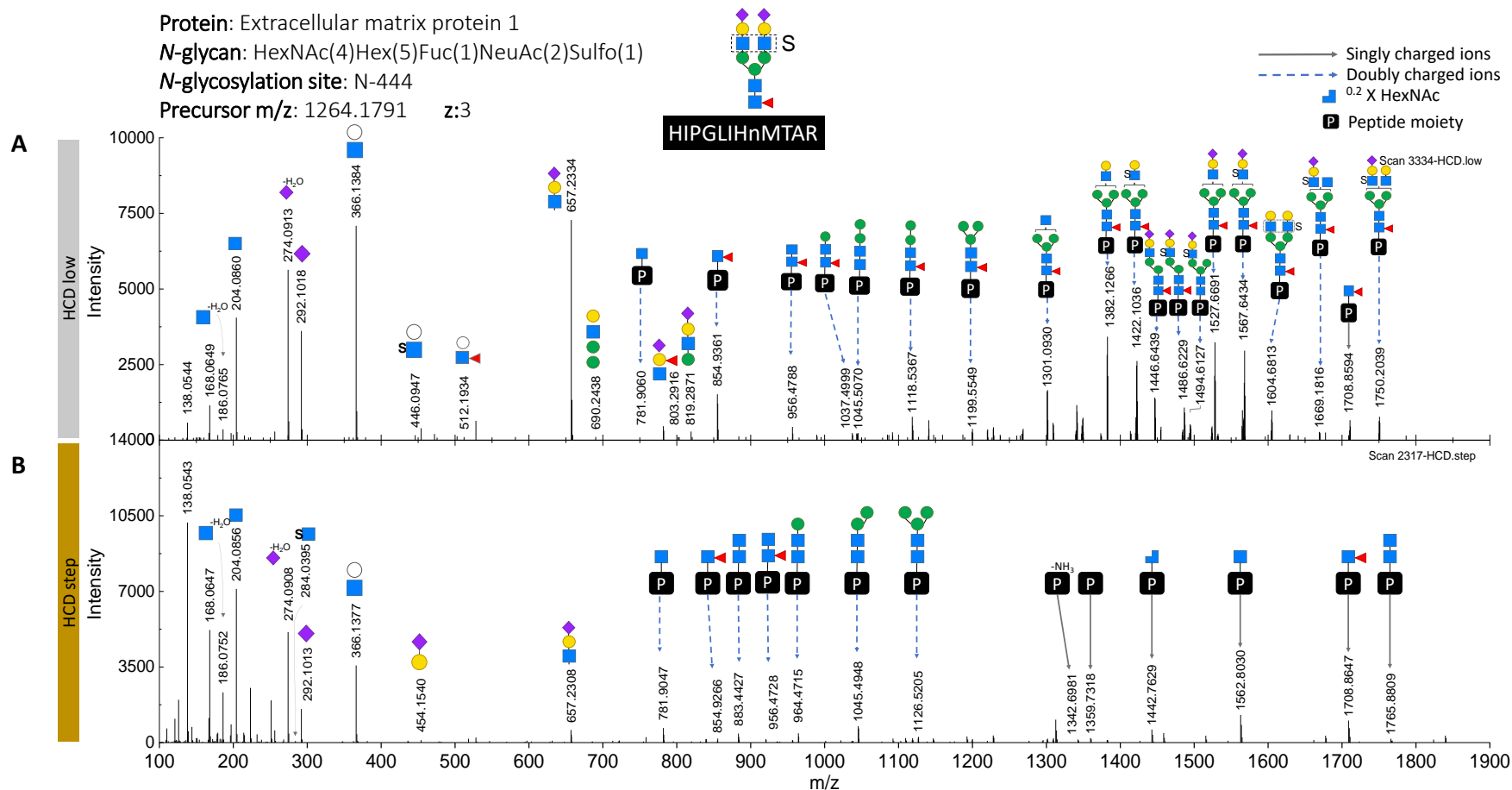


Figure 4.6. Annotation of fragment ions released from sulfated *N*-glycopeptide. (A) HCD.low fragment ion spectrum showing sulfation position. (B) HCD.step fragment ion spectrum showing oxonium ion evidence related to HexNAc sulfation.

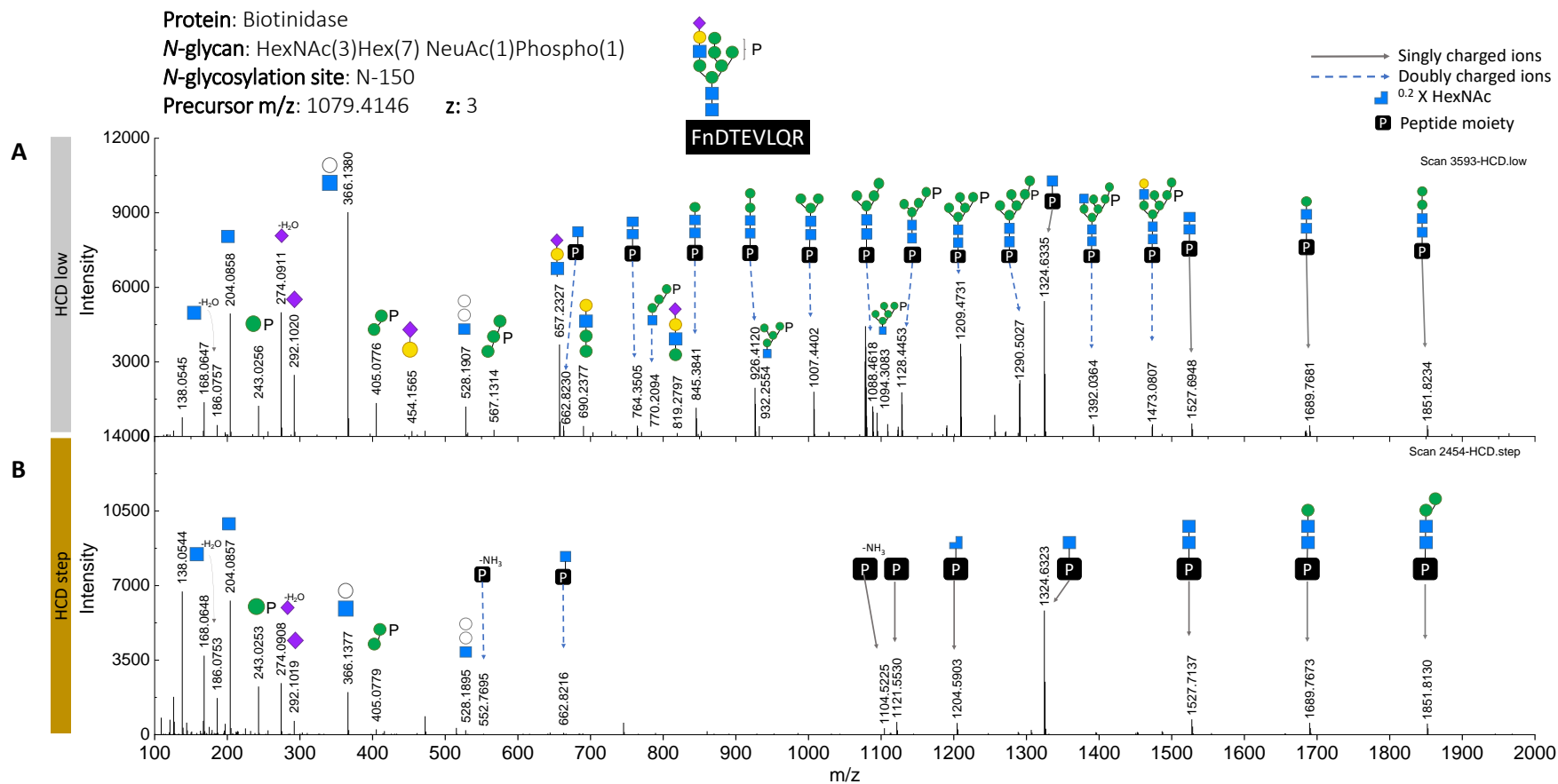


Figure 4.7. Annotation of fragment ions released from phosphorylated *N*-glycopeptide. (A) HCD.low fragment ion spectrum showing phosphorylation position. (B) HCD.step fragment ion spectrum showing oxonium ion evidence related to hexose phosphorylation.

Glycoproteomic analysis of the untreated blood plasma did not allow the detection of phosphorylated nor sulfated *N*-glycopeptides at all (Supplementary Figure A.1.20 Appendix). After analyzing the top 14-HAP depleted sample, one sulfated *N*-glycan attached to N762 from ceruloplasmin and one phosphorylated hybrid *N*-glycan linked to N187 of hemopexin were detected. In contrast, by combining the immunoaffinity depletion of top 14-HAP plus the fractionation by protein molecular weight, the established sample preparation workflow enabled the detection of phosphorylated *N*-glycans in 11 glycoproteins and sulfated *N*-glycans in 54 glycoproteins. Moreover, it allowed the identification of a phosphorylated *N*-glycan in cysteine-rich secretory protein LCCL domain-containing 1, a very-low abundant protein, whose concentration in blood plasma reported by the Human Protein Atlas Database and Protein Abundance Database (PaxDB) is 8.2 pg/mL and 0.033 ppm, respectively [26, 254, 255]. Even though sulfated *N*-glycopeptides tend to lose the sulfated marker ions due to their lability, ten glycoproteins holding HexNAc-sulfated *N*-glycans were confirmed. This demonstrates the power of the preparative workflow for the site-specific identification of modified *N*-glycans.

4.2.6. Rare *N*-glycans in human blood plasma proteins

As displayed in Figure 4.8, three rare *N*-glycan building blocks with the masses 176.0314 Da, 245.0524 Da, and 259.0672 Da were found by *de novo* sequencing of HCD.low MS² spectra of previously incorrect gPSM (see section 3.2.5.2). As described in the following paragraphs, it is hypothesized that the first mass corresponds to glucuronic acid, while the two last possibly resemble other types of acidic sugar. All the identifications with these masses show complex di-antennary monosialylated *N*-glycans. While one antenna is capped by NeuAc, the second one is capped by any rare *N*-glycan building blocks (Figure 4.8). This is supported by the presence of the oxonium ions generated in each case: Hex₁HexNAc₁+176.0314 [M+H]⁺/542.1694, Hex₁HexNAc₁+245.0524 [M+H]⁺/611.1899, and Hex₁HexNAc₁+259.0672 [M+H]⁺/625.2049. Furthermore, site-specific identification of these *N*-glycans revealed their occupancy on N121 from prothrombin (Supplementary Figure A.1.21 Appendix).

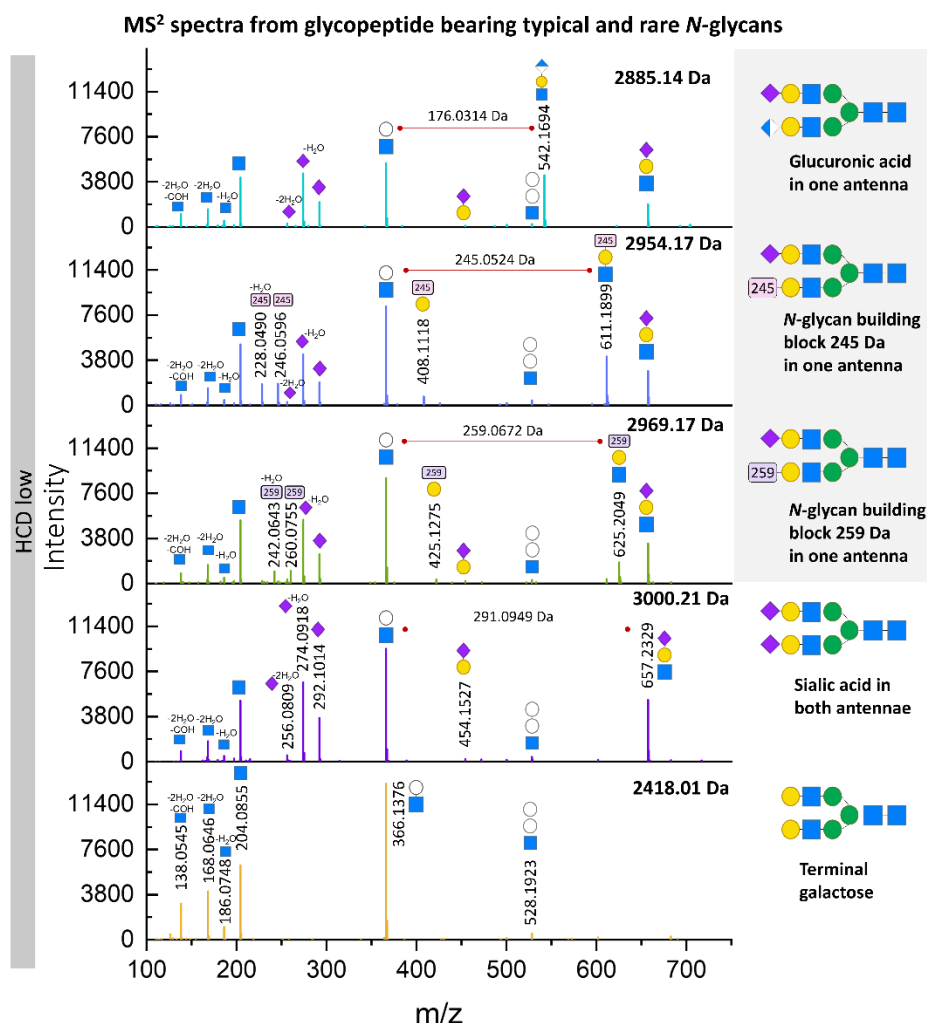


Figure 4.8. Annotation of fragment ions released from typical and rare *N*-glycans attached to N121 (peptide GHVNITR). The rare *N*-glycans appear on a gray background.

The composition of the oxonium ion $\text{Hex}_1\text{HexNAc}_1\text{HexA}_1 [\text{M}+\text{H}]^+$ (542.1716 $[\text{M}+\text{H}]^+$) matches the observed m/z 542.1694 $[\text{M}+\text{H}]^+$ with a mass error of -4.06 ppm. In contrast, if the existence of a methylated hexose instead of a HexA is considered, the m/z of the oxonium ion expected would be 542.2076 $[\text{M}+\text{H}]^+$, which returns an error difference of 67 ppm when comparing with the observed mass. Moreover, the mass of the elemental composition of HexA ($\text{C}_6\text{H}_8\text{O}_6$ after water loss) is 176.0321 Da, and the observed neutral mass is 176.0314 Da. This results in an error of -3.98 ppm between both masses, which favors the hypothesis that hexuronic acid is the first *N*-glycan building block identified. Specifically, it is assumed that the type of HexA present is GlcA since the ion $\text{Hex}_1\text{HexNAc}_1\text{GlcA}_1 [\text{M}+\text{H}]^+$ has been reported as a fragment derived from 3-sulfonated glucuronidated *N*-glycans, holding the glycoepitope HNK-1, common for human brain glycans [251]. Here, the existence of the sulfated HNK-1 glycoepitope is rejected since no fragment

ion nor neutral loss representing GlcA sulfation was detected. To identify other proteins harboring glucuronidated *N*-glycans, a new Byonic™ wildcard search was performed (as detailed in Materials and Methods). As a result, ten glucuronidated *N*-glycopeptides derived from β -2-glycoprotein 1, α -2-HS-glycoprotein, kallikrein, histidine-rich glycoprotein, prothrombin, hemopexin, and complement factor H were identified (Supplementary Table A.1.11 Appendix). No glucuronidated *N*-glycopeptides were detected by analyzing the untreated human blood plasma (Supplementary Figure A.1.20 Appendix). After analyzing the top 14 HAP-depleted samples only the *N*-glycosylation site N453 from hemopexin was found to harbor a glucuronidated complex *N*-glycan. These comparisons demonstrate that the here established sample preparation workflow supports identifying considerably rare *N*-glycopeptides.

The terminal *N*-glycan building blocks 245.0524 and 259.0672 Da have not been described so far. Assuming that these molecules might be a variant of sialic acid (the typical capping sugar), the molecules were searched in PubChem, proposing their hydrated masses (263.20 and 277.23 Da, respectively). The candidates with the highest similarity to Neu5Ac (C₁₁H₁₉NO₉, 309.27 Da [256]) were selected and submitted to CFM-ID 4.0, an online tool to predict fragment ion spectra [257]. The software predicted the fragment ion spectrum using NCE 10 and 20 V (NCE 20 V applied here for HCD.low measurements). After comparing the predicted and the average observed fragment ions, the two candidate molecules shown in Supplementary Figure A.1.21 (Appendix) were selected. The first candidate (C₉H₁₃NO₈, 263.0641 Da [258]) theoretically generates the B ions 246.0608 [M+H]⁺ and 228.0503 [M+H-H₂O]⁺, which are also observed in the MS² spectra that contains the building block “245 Da” (oxonium ions 246.0596 and 228.0490 [M+H]⁺). This results in an error of -4.99 ppm and -5.53 ppm, respectively. The second candidate (C₁₀H₁₅NO₈, 277.0798 Da [259]) produces the oxonium ions 260.0765 [M+H]⁺ and 242.0659 [M+H-H₂O]⁺. These B ions are observed in the MS² spectra where the mass “259 Da” is present (oxonium ions 260.0755 and 242.0643 [M+H]⁺). Comparing the error between the theoretical and observed fragment ions *m/z* for the second candidate, the error results in -3.77 ppm and -6.65 ppm, respectively. The low error supports the hypothesis that the *N*-glycan building blocks found might be two molecules similar to NeuAc, such as the ones presented in Supplementary Figure A.1.22 (in Appendix, molecules drawn in PubChem Sketcher [260]). Both *N*-glycan building blocks were only found on N121 of prothrombin (Supplementary Figure A.1.21 and Supplementary Table A.1.12 Appendix) and have not been reported in this nor other proteins before.

During manual annotation of *N*-glycan structural features in the HCD.low-HBP list, disialo-antennary *N*-glycan structures were found, as shown in Figure 4.9. The surprising presence of one antenna holding two sialic acids was evident by the B ion HexNAc₁Hex₁NeuAc₂ [M+H]⁺/948.3303. Typically, sialic acid is linked to galactose, while the second sialic acid might be linked to this first sialic acid or the antenna-HexNAc. In order to describe this disialo-antennary structure, the MS² HCD.low spectra from two

N-glycopeptides, corresponding to N762 ceruloplasmin and N65 apolipoprotein D, were *de novo* sequenced. In both MS² spectra, the linkage of sialic acid to the HexNAc is demonstrated by the presence of the peptide-HexNAc₃Hex₃NeuAc₁ or the peptide-HexNAc₄Hex₄NeuAc₂ fragment ions ([M+H]²⁺ or [M+H]³⁺). In addition, the fragment ion spectra from apolipoprotein D show the oxonium marker ion HexNAc₁NeuAc₁ [M+H]⁺, supporting the existence of the unexpected HexNAc-NeuAc linkage. Ten proteins presented a disialo-antennary *N*-glycan structure, including ceruloplasmin, α-1-antichymotrypsin, apolipoprotein D, and β-2-glycoprotein 1. All the *N*-glycopeptides presenting disialo-antennary *N*-glycans have the compositions: HexNAc(5)Hex(6)NeuAc(3) or HexNAc(5)Hex(6)Fuc(1)NeuAc(3).

Interestingly, Figure 4.9 also includes the *N*-glycan structure of a disialo-antennary *N*-glycan attached to β-2-glycoprotein 1. This *N*-glycopeptide not only shows one disialo-antenna but also a LacNAc repeat unit on the other antenna. The oxonium mentioned above marker ions are absent in the HCD.step fragment ion spectra annotated by the software and displayed in Supplementary Figure A.1.23 (Appendix). This demonstrates that only by including HCD.low fragmentation more *N*-glycan features are collected to detect relevant glycoepitopes.

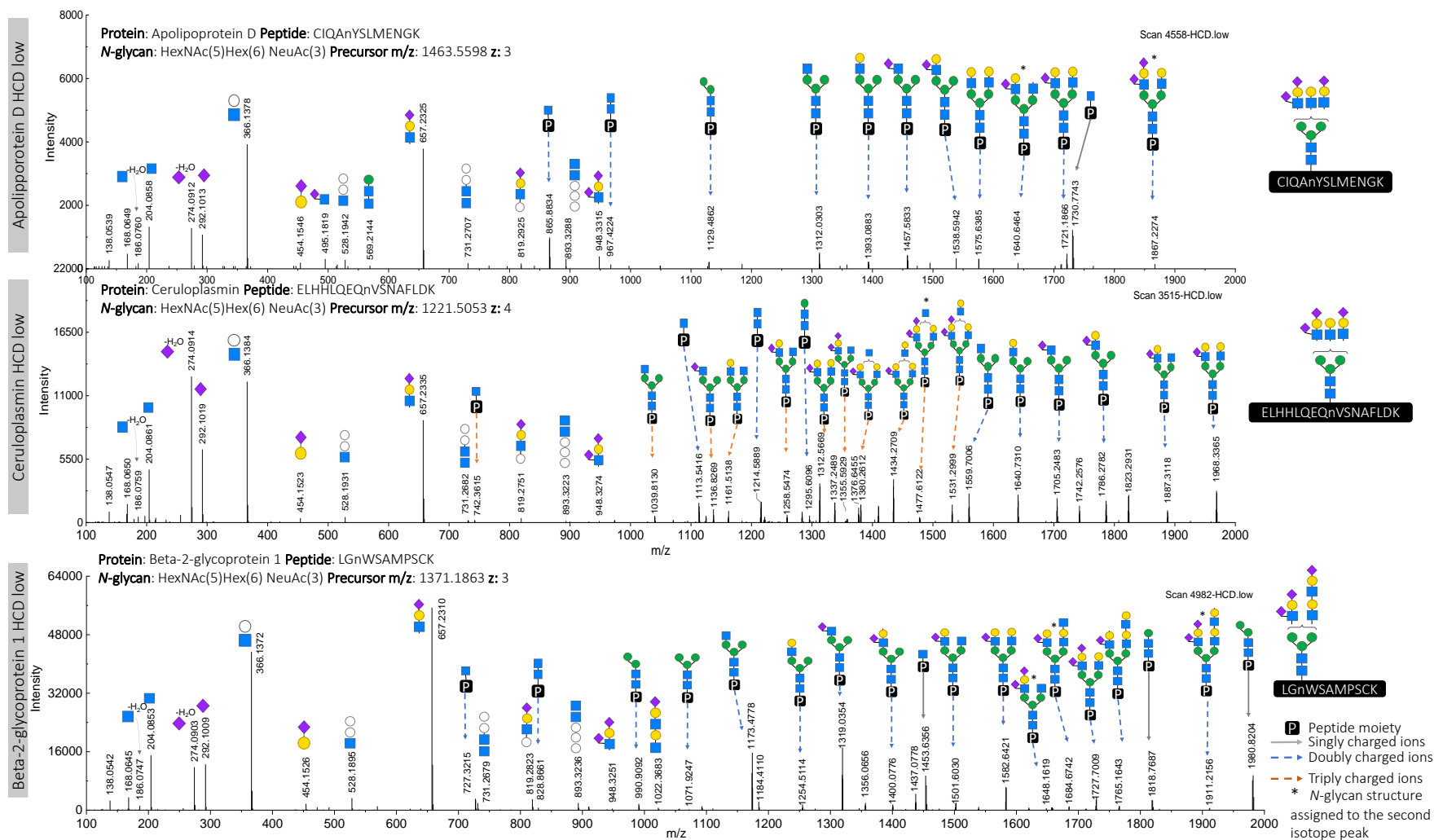


Figure 4.9. *N*-Glycan de novo sequencing of *N*-glycopeptides with disialo-antennary *N*-glycan structure confirming HexNAc-NeuAc linkage.

4.3. Discussion

This part of the thesis describes the established and implemented a workflow for sample preparation and data analysis to conduct an in-depth analysis of the blood plasma intact *N*-glycopeptides. The internal steps in the workflows are connected as follows: 1) conducting blood plasma top 14-HAP depletion followed by a fractionation based on protein molecular weight, tryptic digestion and intact glycopeptide enrichment, 2) measuring glycopeptide enriched fractions twice by nanoLC-ESI OT-OT MS/MS HCD.low and HCD.step fragmentation methods, 3) searching for *N*-glycopeptides using software, 4) manually validating *N*-glycopeptides using the newly developed decision tree, 5) manually annotating structural *N*-glycan features using HCD.low fragment ion spectra and 6) conducting particular glycoproteomic searches to incorporate missed rare *N*-glycan compositions.

4.3.1. Gain of the developed in-depth *N*-glycoproteomic analysis on human blood plasma

The main achievement of this in-depth analysis is the capacity to delve even deeper into the *N*-glycan micro-heterogeneity along with a significantly extended protein concentration range (10^9 – 10^3 pg/mL). Even reaching the detection of low-abundant glycans like phosphorylated *N*-glycans in very low-abundant proteins such as cysteine-rich secretory protein LCCL domain-containing 1 (8.2 pg/mL) [254, 255]. Another example is the reliable identification of sulfated *N*-glycans in middle-abundant proteins (e.g., extracellular matrix protein 1, 0.78 µg/mL [25]) through the detection of sulfated fragment ions, typically hard-to-acquire [185]. The manual validation of all unique *N*-glycopeptide identifications, computed by Byonic™, led to the detection of glucuronidated and other rare *N*-glycans attached to blood plasma proteins expressed in the liver. Finally, the manual annotation of *N*-glycan structural features revealed the advantages of HCD.low spectra for recognizing relevant structures or glycoepitopes. After searching through GlyCosmos, UniCarb, and GlyConnect databases [261-263], it was concluded that there is not yet a site-specific description of the glycoproteins harboring some of the sulfated, phosphorylated, glucuronidated, and disialylated-antenna *N*-glycan structures as identified in our study (Supplementary Tables A.1.13 and A.1.14 Appendix) [154, 264-267]. Then, by performing a proteomic analysis on all fractions after *N*-glycopeptide enrichment (FDR <1%), it was demonstrated an expanded detection limit of blood plasma proteins within a concentration range from $1 \cdot 10^9$ down to $1 \cdot 10^3$ pg/mL. This evaluation not only demonstrates that our workflow is comparable to other in-depth proteomic workflows [111, 113] but also that it has significant advantages for glycoproteomics in blood plasma. In contrast, the study from Wessels *et al.* presented a strategy for diagnosing congenital disorders of glycosylation (CDG) via the direct glycoproteomic analysis of untreated blood plasma [268]. Their method resulted in the site-specific profile of 34 proteins, which satisfied the evaluation of the CDG cases selected. Our workflow offers an alternative

for evaluating site-specific profiles of low-abundant proteins and other *N*-glycan types, like rare *N*-glycan structures that might be involved in different CDGs. So far, structural *N*-glycan details require glycomics or exoglycosidases and lectins. However, here it was shown a new avenue for moving towards structural glycoproteomics. This study aimed to address the limitations and pitfalls caused by the blind spots existing during an *N*-glycoproteomic analysis. By acknowledging these aspects, glycoproteomic bioinformatics tools can be refined to obtain more accurate and comprehensive results.

4.3.2. Comparison of the performance of this workflow to other studies

Recently, many studies revealed substantial information about the blood plasma *N*-glycoproteome based on the analysis of intact *N*-glycopeptides [111-113]. These analyses not only applied multi-step sample preparation workflows but also powerful bioinformatics tools. The most extensive study was performed by Shu *et al.*, achieving the identification of 1,036 *N*-glycosites containing 738 *N*-glycans, resulting in 22,677 unique *N*-glycopeptides derived from 526 glycoproteins using pMatchGlyco software (FDR 1%) [111]. Our study obtained 7,867 *N*-glycopeptide identifications (no decoys, no cuts on FDR), from which 1,929 were manually validated as “True” identifications, revealing 942 different *N*-glycosites and 805 human glycoproteins. The higher number of *N*-glycopeptide identifications of Shu *et al.* might result from a library of de-*N*-glycosylated peptide identifications created by cumulative searches, including semi-tryptic digestion and 16 variable peptide modifications. In addition, they explored atypical *N*-glycosylation consensus sequences (*N*-X-S/T/V/C, X≠P) and an extensive *N*-glycan database (739 *N*-glycans). In contrast, our study resulted from one standard search applying full specific tryptic digestion, four variable peptide modifications, common *N*-glycosylation consensus sequence (*N*-X-S/T, X≠P), and 288 *N*-glycan compositions. Shu *et al.* used a pooled serum sample from 50 healthy individuals, compared to the normal control plasma from VisuConTM-F (20 donor plasma pool) used in this project. The higher number of donors might increase the chance of accumulating the proteins secreted in blood plasma in different abundances. Additionally, the authors included the high-abundant protein-depleted fraction in their study. In comparison, the analysis conducted here resulted from one blood plasma sample from which the top 14 high-abundant proteins were depleted and later discarded using a single-use immunoaffinity column. Analysis of the high-abundant protein fraction was not in the scope of our study since *N*-glycosylation of these proteins is well studied. Another important limitation in the here developed study was the relatively long measurement time of the instrument used for acquiring data (Orbitrap Elite-Velos, scan speed 4 Hz), compared to Orbitrap Q Exactive mass spectrometer (with three times faster scan speed of 12 Hz) used by Shu *et al.* The significant difference between the performance of both instruments was demonstrated by Sun *et al.* [133].

Software and parameters for the search are other factors that can lead to different results regarding the interpretation of *N*-glycopeptides. This is shown in the study from Kawahara *et al.*, where two glycoproteomic spectra files derived from human serum were provided to 22 expert groups in glycoproteomics to evaluate the impact of applying different bioinformatics strategies on the identification of intact *O*- and *N*-glycopeptides [148]. The results showed an enormous variability of glycopeptide identifications, glycoproteins, and glycan compositions. The study attributed this inconsistency mainly to the filters applied after the search, such as score threshold or FDR cut. Here, it was excluded a post-search filter interference by disallowing decoys and FDR cuts. Instead, a manual validation after the *N*-glycopeptide search was conducted, acknowledging that this approach could result in incorrect identifications.

Nevertheless, the scope of the study conducted here was to identify hitches and opportunities given during intact *N*-glycopeptide analysis, especially for incorrect *N*-glycopeptide identifications with high quality-score and ambiguous gPSM. All the identifications were manually scrutinized, evidence supporting the trueness of the identification was annotated, and incorrect matches were revised. Thus, a validation subcategory related to the variability observed by Kawahara *et al.* was the “True evidence with alternatives” subcategory. This subcategory lists *N*-glycopeptides whose alternative peptide matches are named in the column “comments” of the table HCD.step-HBP list in Supplementary Table A.1.6 (Appendix). In the dataset generated in this study, 35% of the “True” *N*-glycopeptides are classified within this subcategory, and two common features were observed: 1) a low number of b and y ions in the MS² spectra and 2) a low abundance of these observed proteins in blood plasma. The first issue relates to low MS² spectra quality, associated with several factors such as precursor ions with poor ionization efficiency, fragment ion losses, and low-abundant precursor ions. The second issue depends on the protein concentration distribution of the sample, where low-abundant proteins are reflected as low-abundant precursor ions in the final spectra. Regarding the second feature, Kreimer *et al.* propose implementing algorithms designed for smart selection and acquisition of the typically suppressed precursor ions to reduce the proportion of spectra with poor-quality [269]. Then, these algorithms would reduce the proportion of spectra that lead to the variability of glycopeptide identifications by improving spectra quality during the MS/MS measurement.

4.3.3. Observations on the validation of *N*-glycopeptides with multiple fucoses

Another source of variability among glycoproteomic analyses is fucosylation. After manual validation, a high frequency of identifications presenting multiple fucose moieties without any fucose-related fragment ion was observed. The occurrence is more elevated in di- tri- or tetra-antennary *N*-glycopeptides with an incomplete number of capping sialic acid in the antennae. Kawahara *et al.* reported a similar observation,

where the high frequency of *N*-glycopeptide identifications with multiple fucose moieties, reported by many participants, did not correlate with the results obtained from typical *N*-glycomic analysis of human blood plasma. It was observed that an incorrect detection of the isotopic pattern for the corresponding precursor ion causes the assignment of multiple fucose moieties instead of one sialic acid. Lee *et al.* also observed this problem and reported that it might be influenced by the precursor-picking default settings [151]. Hence, special attention must be given to this phenomenon since this problem could also induce the incorrect assignation of other not fucose-related quasi-isobaric *N*-glycan masses, such as HexNAc(5)Hex(6)Fuc(1)NeuAc(3) / 3007.0580 Da and HexNAc(7)Hex(6)NeuGc(2) / 3008.0532 Da.

Determination of the fucose position on the *N*-glycan is an additional difficulty when deciphering the structure of a fucosylated *N*-glycan. Hexose rearrangement is a reaction often found in MS where an internal hexose migrates to a different position within the *N*-glycan, producing “false” *N*-glycan structures [270]. Fucose rearrangement occurs in the gas phase, resulting in a fucose transfer between two antennae or between core and – most likely – α 6-Man-linked antenna (due to its flexibility) [253]. Consequently, this reaction generates misleading fragment ions such as HexNAc₁Hex₁Fuc₁ in the MS² HCD.step spectra of only core fucosylated *N*-glycopeptides. Acs *et al.* showed that core-fucose linkage is robust at high NCE energies. However, Acs *et al.* and Wuhler *et al.*, also acknowledge that a transfer of fucose from the antenna to the core can be induced since they observed the Y ion (peptide-HexNAc₂Hex₃Fuc₁ [M+H]⁺). In the study developed here, that Y ion was not used for confirming core fucosylation but primarily the peptide-HexNAc₁Fuc₁ [M+H]⁺ Y ion. Diverse studies reported that core and antenna fucose-linkage stability improves by using NCE 20 instead of higher collisional energies, even though fragments from fucose rearrangement might still be produced in lower abundance [118, 271]. On the one hand, in agreement with these studies, it was observed the ambiguous generation of fucose fragment ions, e.g., the detection of both peptide+HexNAc₁Fuc₁ (Y-ion) and HexNAc₁Hex₁Fuc₁ [M+H]⁺ in the MS² spectra of monofucosylated *N*-glycopeptides. It is assumed that core fucosylation would most likely lead to the generation of an antenna-marker ion (instead of the opposite rearrangement). On the other hand, the adequate chromatographic separation of antenna fucosylated and core-fucosylated labeled *N*-glycans on a C18 column, while it has been demonstrated [272], might not be achieved for all the glycopeptides in a complex sample [128, 273], such as blood plasma, and the overlap of structural isomers can occur. Therefore, coelution of isomers (antenna- and core-fucosylated peptides) cannot be ruled out as a possible situation explaining the presence of both ions. Thus, the monofucosylated gPSMs showcasing both ions were categorized as “core- or antenna-fucosylation”.

With regard to the reliability of the described multiple-fucosylated *N*-glycan structures, it is necessary to check many details, such as Y ions, neutral losses, and the MS¹ isotopic pattern distribution, using the

spectra acquired at low collisional energies. For this, manual annotation is possible but inefficient if the validation involves an increased number of gPSMs containing fucosylated *N*-glycans.

4.3.4. Relevance of *N*-glycan structure elucidation

Over the last few years, particular emphasis has been given towards describing *N*-glycan structural information and their localization within the protein [129]. Even though glycomic analysis has successfully added to the description of structural groups in *N*-glycans (diLacNAc unit, sialyl Lewis X, phosphorylation, etc.), only MS allows to determine the original position of such *N*-glycans precisely. Intending to integrate *N*-glycan structural information using MS, Shen *et al.* conducted an *N*-glycoproteomic analysis of mouse brain tissue using the new software “StrucGP”. By acquiring and combining two complementary fragmentation energies (HCD.low for glycan and HCD.step for peptide moiety), this software could interpret *N*-glycan structural groups on intact *N*-glycopeptides [145]. Similarly, information from two MS² analyses (HCD.step and HCD.low) was coupled and manually annotated the oxonium marker ions that support identifying special *N*-glycan structural features and modifications, as previously described by our group [118]. From observing fragment ions, *N*-glycan isoforms such as bisecting *N*-glycans and repeated LacNAc structure were identified. These identifications were searched in GlyConnect to identify their corresponding reported information [262, 264]. The bisecting *N*-glycans observed in immunoglobulins proteins and plasma protease C1 inhibitors agree with information reported in GlyConnect [262, 264]. However, no reports were found for the rest of the identified *N*-glycopeptides with bisecting GlcNAc (Supplementary Table A.1.14 Appendix). Even though the *N*-glycosylation sites harboring LacNAc structure have been reported in terms of *N*-glycan composition, evidence of sialylated diLacNAc structure was not described before for the *N*-glycosylation sites identified in this work. In addition, an interesting disialo-antennary *N*-glycan was observed in the N253 of β -2-glycoprotein 1. A similar *N*-glycan is reported for the site N162 of this protein [264].

The disialylated-antenna structure found in our study might resemble the epitope disialyl Lewis C NeuAc α 2-3Gal β 1-3(NeuAc α 2-6)GlcNAc, reported in α -2-HS-glycoprotein from *bos taurus* [274]. More experiments are necessary to achieve sialic acid linkage elucidation. Based on previous reports, sialic acid linkage (α 2,3 and α 2,6) determination, on intact glycopeptides, can be not only achieved through ion mobility [275, 276], but also by triggering MS³ fragmentation of the HexNAc₁Hex₁NeuAc₁ oxonium ion [277, 278]. Hence, integrating MS³ fragmentation of the HexNAc₁Hex₁NeuAc₁ and HexNAc₁Hex₁NeuAc₂ oxonium ions, as part of the LC-MS/MS method, can support the structural description of this rare *N*-glycan, to a certain extent. Saraswat *et al.* also identified sialic acid linkage to a HexNAc in an *N*-glycan attached to clusterin protein [112]. At the same time, they also found nine other proteins presenting *N*-glycans with linkage between two sialic acids. However, no *N*-glycosylation sites were in common

between Saraswat *et al.* and this study. Only the protein α -1-antichymotrypsin was identified in both studies with a disialylated antennary group in different *N*-glycosites [112]. The NeuAc α 2-3Gal β 1-3(NeuAc α 2-6)GlcNAc epitope was also identified in isolated proteins (transferrin, α -1-antitrypsin, Immunoglobulin G and α -1-acid glycoprotein) and the serum *N*-glycome from a patient with an atypical neurological phenotype and liver adenomatosis, by Sturiale *et al.* [4]. The production of GlcNAc-NeuAc linkage in human *N*-glycans has been studied previously [279, 280]. Although a recent study proposes that NeuAc rearrangement at low NCE energies generates low levels of GlcNAc-NeuAc oxonium ion, additional ions can support the presence of the here observed disialo-antennary glycoepitope [281]. However, our work proposes the existence of this disialo-antennary *N*-glycan in the low-abundant normal blood plasma *N*-glycoproteome, providing evidence from oxonium ions but also from Y ions.

4.3.5. Identification of glycopeptides bearing sulfated and phosphorylated *N*-glycans

Sulfation and phosphorylation are two post-glycosylational modifications that add a negative charge to the *N*-glycan. The *N*-glycopeptides holding this type of glycan modifications (as well as sialylated *N*-glycans) are better detected by methods that enrich anionic molecules [155, 282]. Nevertheless, the study developed hereby showed that, while none of these modified *N*-glycopeptides were detected via the analysis of the untreated human blood plasma, their detection is possible after applying a multi-step sample preparation workflow. *N*-Glycopeptides featuring sulfated and phosphorylated *N*-glycans were detected for proteins with reported concentrations ranging from $4.17 \cdot 10^8$ down to $1.55 \cdot 10^5$ pg/mL (PPD, plasma proteome database) [25]. Results show that sulfated *N*-glycopeptides are more frequently present than phosphorylated *N*-glycopeptides. However, *N*-glycopeptide identifications holding phosphorylated *N*-glycans always showed corresponding oxonium marker ions –which was not the case for sulfated *N*-glycopeptides. It seems that phosphate-containing fragment ions are more stable than sulfate-containing fragment ions. This observation is in agreement with Zhang *et al.*, who found that the abundance of Hex₁Phospho₁ [M-H₂O+H]⁺ oxonium ions was higher than Hex₁Sulfo₁ [M+H]⁺ using low fragmentation energy [283]. From the sulfated *N*-glycopeptide identifications observed, only 17% showed the oxonium ion HexNAc₁Sulfo₁ [M+H]⁺ within their MS² spectra, probably due to the short lifetime of sulfated fragment ions in MS analysis [185]. Recent glycomic work of our group confirmed the presence of HexNAc sulfation in *N*-glycans released from a human serum immunoglobulin A (IgA) [44, 45]. Cajic *et al.* established a versatile glycomic workflow that enables the isolation of any *N*-glycan of interest via HILIC-HPLC-FLD (by means of a removable fluorescent dye), followed by multiple analysis (MALDI-TOF-MS and xCGE-LIF) [45]. Employing this glycomic workflow, Cajic *et al.* identified a sulfated *N*-glycan released from human serum IgA, for which Chuzel *et al.* further determined the presence of GlcNAc-6-SO₄ using a highly specific sulfatase found via functional metagenomics [44].

Spurious sulfated *N*-glycopeptide identifications might be caused by the assignment of sulfation plus two HexNAc instead of three Hexoses (Hex₃=486.1585 Da and HexNAc₂Sulfo₁=486.1156 Da). Shu *et al.* detected phosphorylated and sulfated *N*-glycans in many blood plasma proteins, including some also found here. In contrast to our studies, all the sulfated *N*-glycans identified in their study show hexose sulfation instead of a HexNAc sulfation [111]. A previous study showed that *N*-glycans holding a sulfated-galactose (like Gal-3-sulfate) do not generate the ion Hex₁Sulfo₁ [M+H]⁺; however these sulfated *N*-glycans generate HexNAc₁Hex₁Sulfo₁ [M+H]⁺ (scarcely) and/or a sulfate neutral loss between the Y ion including galactose and the consecutive sulfated-Gal Y ion [245]. This suggests that some of our sulfated *N*-glycopeptide identifications lacking sulfated fragment ions might correspond to galactose-sulfated *N*-glycopeptides. However, it is more difficult to collect this evidence for all identifications since it is required to annotate manually the Y ions observed in each HCD.low MS² spectra. During the manual validation in the HCD.step MS² spectra, no HexNAc₁Hex₁Sulfo₁ [M+H]⁺ oxonium ions were detected but only HexNAc₁Sulfo₁ [M+H]⁺. After *de novo* sequencing in HCD.low MS² spectra, the HexNAc sulfation was confirmed in the antennae, most likely corresponding to sulfated-6-GlcNAc (Figure 4.6). The majority of the sulfated *N*-glycopeptides found here feature diantennary, mono- or di-sialylated, and sometimes core-fucosylated *N*-glycans with LacNAc extensions, resembling the LacNAc sulfated glycoepitope: 6-sulfo sialyl Lewis X in its defucosylated form. The *N*-glycopeptide identifications containing confirmed-sulfated *N*-glycans derive from glycoproteins involved in blood coagulation, fibrinolysis, hemostasis, inflammatory response, mineral balance, osteogenesis, complement pathway, apoptosis, and innate immunity [284].

LacNAc sulfated *N*-glycans have various implications in immunology. On the one hand, 6-sulfo sialyl Lewis X (with GlcNAc-6-SO₃) is a ligand for L-selectin, a cell adhesion molecule for tethering and trafficking of lymphocytes through the peripheral nodes [89]. On the other hand, the structural isomer, 6'-sulfo sialyl Lewis X (with Gal-6-SO₃), is a primary ligand of Siglec-8. The cross-linkage between 6'-sulfo sialyl Lewis X and Siglec-8 in mast cells induces histamine and prostaglandin D₂, whereas it induces apoptosis in eosinophils [246, 285]. A glycomic analysis performed by Yamada *et al.* on the serum of patients with pancreatic cancer showed that the abundance of sulfated *N*-glycans is increased compared to healthy controls [29]. The study also showed that the phosphorylated *N*-glycans remained stable, stressing the importance of reliably discriminating both *N*-glycan modifications in pathophysiology.

Sleat *et al.* detected for the first time Man-6-P in high-abundant plasma glycoproteins [248]. They calculated the relative fraction of proteins with phosphorylated *N*-glycans in blood plasma and compared it with the fraction of proteins with phosphorylated *N*-glycans from the lysosome, concluding that phosphorylated *N*-glycans in blood plasma proteins exist rather as traces. It is known that Man-6-P is recognized by transporters (Man-6-P receptors) that carry the Man-6-P modified glycoprotein to the lysosomes [248]. This receptor, expressed in all human cells and tissues, is not only present intracellularly

in Golgi apparatus and endosomes but also extracellularly in cell membranes [249]. In the extracellular context, Man-6-P receptors like CD222 are involved in protein internalization, trafficking, lysosomal biogenesis, apoptosis, cell migration, and regulation of cell growth [249]. This suggests that interactions between mannose-phosphorylated *N*-glycans and binding molecules like CD222 (a receptor from the P-lectin family) are essential in the physiology [67, 249].

4.3.6. Identification rare *N*-glycans in incorrectly-assigned gPSMs

Another *N*-glycan modification not regularly explored in the clinical context is glucuronidation. *N*-glycopeptides containing glucuronic acid were observed after the manual curation of incorrectly assigned *N*-glycopeptide identifications. The incorrect matches might have resulted from the assignation of two NeuAc instead of two HexNAc plus one GlcA residue, which have in total the same atomic composition ($2\text{NeuAc} = 2[\text{C}_{11}\text{H}_{17}\text{O}_8\text{N}] = \text{C}_{22}\text{H}_{34}\text{O}_{16}\text{N}_2$, and $2\text{HexNAc} + 1\text{GlcA} = 2[\text{C}_8\text{H}_{13}\text{O}_5\text{N}] + 1[\text{C}_6\text{H}_8\text{O}_6] = \text{C}_{22}\text{H}_{34}\text{O}_{16}\text{N}_2 = 582.1908 \text{ Da}$). Huffman *et al.* observed glucuronidated *N*-glycans in the blood plasma *N*-glycome of different European populations [153]. Yamada *et al.* reported a reduced relative abundance of glucuronidated *N*-glycans from the serum of patients with pancreatic cancer compared to healthy controls [29]. Sulfated glucuronic acid is more common in brain *N*-glycans as a key component of the glycoepitope HNK-1 ($\text{SO}_3\text{-3GlcA}\beta\text{1-3Gal}\beta\text{1-4GlcNAc}$). HNK-1 influences neuronal functions like adhesion, cell recognition, migration, preferential motor-re-innervation, synaptic plasticity, and for post-trauma regeneration in the peripheral and central nervous system [286, 287]. Laminins and cadherin-2 are HNK-1 binding proteins [287]. Nevertheless, it is unclear if this is relevant to the biological role of the non-sulfated HNK-1.

Two new *N*-glycan building blocks were identified with the masses 245.0524 and 259.0672 Da attached to complex di-antennary monosialylated *N*-glycans. These masses are unrelated to the also atypical ketodeoxynononic acid identified by Wang *et al.* as a capping sugar in *N*-glycans from Human Prostate-Specific Antigen purified from seminal fluid [288]. None residue resembling such *N*-glycan building blocks in humans were identified after literature search. A limitation of this observation is the lack of an orthogonal method to describe the molecule structure. Interestingly, both rare-*N*-glycan building blocks were found in prothrombin (glycosylation site N121). This *N*-glycosylation site corresponds to prothrombin fragment region 1 (kringle-1), essential for calcium-mediated membrane-surface binding [289].

Bioinformatics tools able to provide a site-specific *N*-glycan description from the analysis of intact glycopeptides emerged 12 years ago [146]. Since then, software design has moved towards improving gPSM quality (pGlyco 2.0), visualization of results (pGlyco 2.0, glyXtoolMS, GPSeeker), and annotation of *N*-glycopeptide structural features (GPSeeker, glyXtoolMS, StrucGP) [132, 145, 149, 243]. Nonetheless,

challenges that prevent the accurate analysis of an *N*-glycoproteome are the lack of comprehensiveness, ambiguity on peptide composition or glycan structure, missing structural information, and false-positives [129]. Consequently, the detection of *N*-glycopeptides potentially useful as biomarkers or for therapy is hampered. Even though manual validation is an alternative for detecting pitfalls after a glycoproteomic search, it has discouraging aspects, such as the high time requirement and effort needed, plus the bias caused by each data reviewer. For future developments, it would be appropriate for *N*-glycoproteomic software to include an *N*-glycan diagnostic step plus the integration of *N*-glycan structural features during glycopeptide spectra matching. Missing *N*-glycan compositions, lack of information, or incorrect parameters are barriers against fully explaining an *N*-glycoproteomic spectra input. Our blood plasma *N*-glycoproteomic analysis aimed to find *N*-glycopeptides eclipsed between the obscure niche of the low-abundant *N*-glycoproteome and the MS² spectra neglected due to insufficient search parameters. After manual validation, the categories “Uncertain change *N*-glycan and peptide” and “Uncertain glycopeptide” addressed that 51% of total gPSMs are still unexplained, for a greater part, due to the low quality of the fragment ion spectra. Nonetheless, it was possible to identify several previously unknown and rare *N*-glycans plus various atypical *N*-glycan structures that are potentially useful for the design of biotherapeutics, clinical diagnostic (biomarker discovery), or exploration of *N*-glycan-protein interactions and functions.

4.4. Summary

To understand the implications of protein glycosylation for clinical diagnostics and production of biopharmaceuticals, innovative glycoproteomic technologies are required. Recently, significant advances were made, particularly toward structure-focused (*N*-glyco-)proteomic analyses. The mass spectrometric analysis of intact *N*-glycopeptides using stepped collision fragmentation and glycan oxonium ion profiling enables reliable discrimination between different *N*-glycan structures. Still, there are weaknesses that current *N*-glycoproteomic approaches must overcome: 1) handling of incorrect identifications, 2) identification of rare and modified *N*-glycans, and 3) insufficient glycoproteomic coverage, especially of complex samples. To address these shortcomings, an innovative *N*-glycoproteomic workflow that aims at providing comprehensive site-specific and structural *N*-glycoproteomic data on human blood plasma glycoproteins was established. The workflow features protein depletion plus a fractionation strategy and the use of high-resolution mass spectrometry with stepped collision fragmentation. Furthermore, by including a new decision tree procedure developed for data validation, it was possible to significantly improve the description of the *N*-glycan micro-heterogeneity. This advanced data analysis workflow allows the reliable differentiation of ambiguous *N*-glycan structures like antenna- versus core-fucosylation plus the modified and rare *N*-glycans such as sulfated and glucuronidated ones. This workflow, enabled the

detection of human blood plasma glycoproteins with reported concentration within the ng/mL level. A total of 1,929 *N*-glycopeptides and 942 *N*-glycosites derived from 805 human middle- to low-abundant glycoproteins were identified. Overall, the presented workflow holds great potential to improve our understanding of protein glycosylation and to foster the discovery of blood plasma biomarkers.

5 Site-specific identification of sulfated and *O*-acetylated *N*-glycans in human serum IgA

Please note that parts of this section are taken from the original publication Zuniga-Banuelos *et al.* 2024 and 2025b [227, 229].

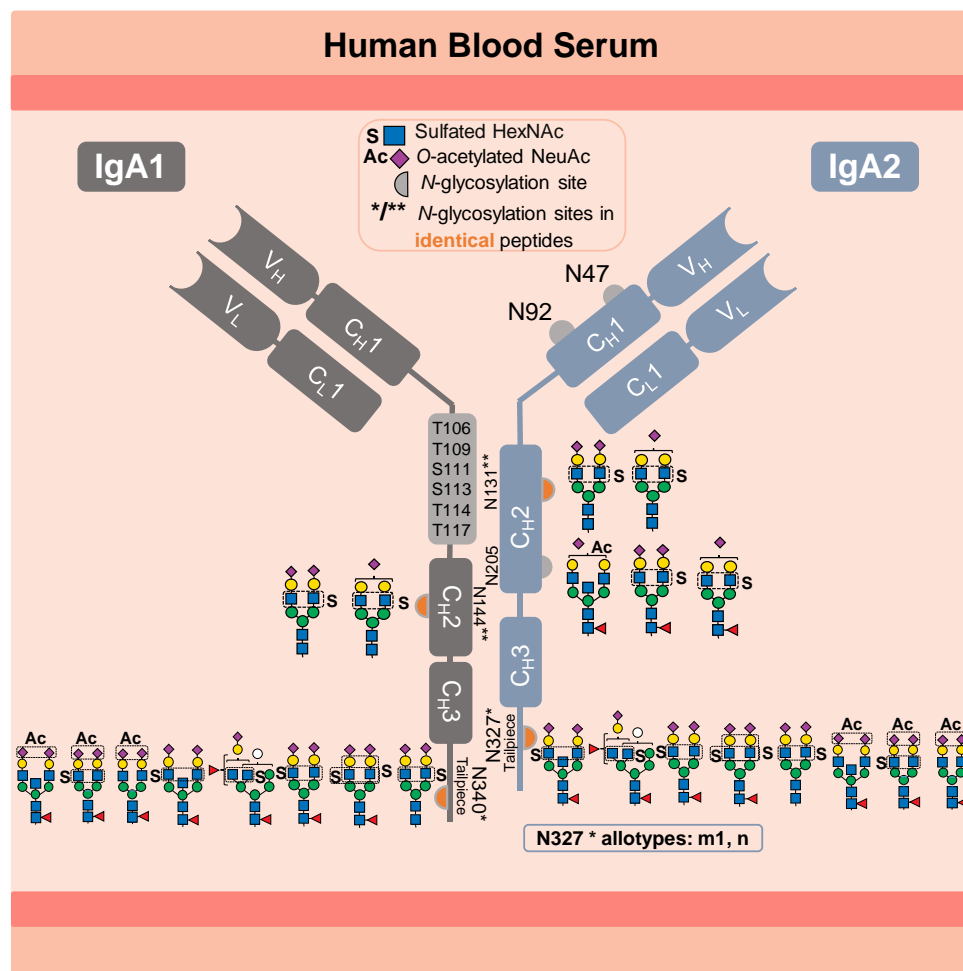


Figure 5.1. Description of diverse sulfated *N*-glycans identified by us at the *N*-glycosylation sites of human serum IgA1 and IgA2. The numbers after “capital N” stands for the amino acids bearing *N*-glycosylation sites. The numbers after “capital S or T” represent the amino acids bearing *O*-glycosylation sites.

5.1. Introduction

As described in section 2.3.2., human IgA is a more heterogeneous immunoglobulin than human IgG. IgA not only displays different structural conformations, subclasses, polymorphisms, and *O*-glycosylation but also has more *N*-glycosylation sites [178, 179]. An overview of IgA glycosylation sites is shown in Figure 5.1. The *N*-glycosylation profile varies between IgA1 and IgA2 subclasses (described in section 2.3.2.), and *N*-glycosylation sites. A lectin blot analysis of each serum IgA subclass revealed that IgA2 bears fewer sialylated, galactosylated, and bisected *N*-glycans and more hybrid- and oligomannose-type *N*-glycans than IgA1 [184]. Furthermore, these features seem critical for the pro-inflammatory effect of and the anti-inflammatory modulatory role of IgA2 and IgA1, respectively [184]. In addition, several studies have used bottom-up mass spectrometry analyses to obtain a site-specific description of the serum IgA *N*-glycosylation [7, 110, 152, 183, 184, 290]. However, the *N*-glycans observed at the N144-IgA1 and N131-IgA2 sites (in the C_H2 domain), as well as the tailpiece sites N340-IgA1 and N327-IgA2_{m1/n} (in the C_H3 domain), are reported to both IgA subclasses due to the identical amino acid sequences surrounding these sites. Yet, significant differences between the *N*-glycosylation profile of both C_H2 and C_H3 domains were observed. The C_H2 domain features mostly non-fucosylated hybrid-type *N*-glycans and di-antennary complex-type *N*-glycans with terminal galactose or sialic acid. In contrast, the tailpiece typically bears core fucosylated di- or multi-antennary complex-type *N*-glycans variably sialylated and bisected, as well as oligomannose-type *N*-glycans.

N-glycosylation influences IgA functions (Figure 2.8). Some studies showed that *N*-glycans attached to the sites in the C_H2 domain and the tailpiece (IgA1 Fc *N*-glycans) are irrelevant for binding to the FcαRI receptor [290, 291]. However, other studies indicate that IgA1 Fc *N*-glycans are relevant for binding of other receptors involved in anti-inflammatory mechanisms [174, 184]. Closer investigations on the *N*-glycosylation at the tailpiece have confirmed its high impact on different IgA functions. First, *N*-glycosylation at the tailpiece is critical for binding to complement C3 protein [188]. Second, it was shown to be essential for modulating dimer formation in IgA1 [188]. Third, *N*-glycosylation at the tailpiece enhances the IgA anti-viral neutralizing activity by exposing sialylated *N*-glycans and interacting with sialic-acid-binding viral proteins, like influenza hemagglutinin and neuraminidase [172]. Fourth, the tailpiece site affects the serum half-life of IgA1 [189]. Rifai *et al.* showed that the clearance of IgA1 from serum is slower when *N*-glycans at the tailpiece are removed, whereas it remained unchanged when *O*-glycans are removed [189]. Although the function of IgA2 *N*-glycans on the C_H1 domain (Fab region) remains unclear, Rifai *et al.* suggested that the additional *N*-glycosylation sites on IgA2 (N47 and N92 in the C_H1 and N205 in C_H2) alter the rate of IgA2 clearance from the blood. These findings suggest to further characterize IgA

N-glycosylation in terms of its micro-heterogeneity (describing the site-specific *N*-glycan variability) for elucidating its role in the mechanisms modulating critical effector functions.

In addition to the complexity caused by glycosylation, IgA has a prominent abundance of *N*-glycans modified by sulfation. This was detected for the first time in 1999 by Boisgard *et al.* in IgA from mammary glands of rabbits, but also from to other species and tissues [28]. However, sulfation could be confirmed for human serum IgA only 20 years later by in-depth glycomic analyses [28, 44, 45]. Identification of sulfated *N*-glycans is a challenging task. In particular, in other glycomic and glycoproteomic studies conducted on IgA, sulfated *N*-glycans never appeared in the identification lists [7, 110, 152, 183, 184, 290, 292]. Recently, Cajic *et al.* and Chuzel *et al.* confirmed the presence of the sulfated *N*-glycan FA2G2S2-SO₄ in human serum IgA [44, 45]. Chuzel *et al.* demonstrated that the sulfate group was linked to the 6-carbon of GlcNAc (GlcNAc-6-SO₄) by the application of a highly-specific sulfatase in combination with a method developed by Cajic *et al.* [44]. In this method, the sulfated *N*-glycan FA2G2S2-SO₄ was released upon cleavage of IgA *N*-glycans and all labeled with the removable fluorescent label 9-fluorenylmethyl chloroformate (Fmoc) [45]. The Fmoc-labeled sulfated *N*-glycan was isolated by HILIC-HPLC-FLD fractionation, identified, and characterized employing two orthogonal approaches: MALDI-TOF-MS and a dedicated enzymatic approach combined with xCGE-LIF *N*-glycan analysis [45]. Cajic *et al.* demonstrated that the newly discovered sulfated *N*-glycan, with the composition is HexNAc₄Hex₅Fuc₁NeuAc₂Sulfo₁, presents α2-6-Neu5Ac on both antennae and GlcNAc-6-SO₄ on the α1-3-Man arm. However, the glycosylation sites, could not be determined yet [45].

Alagesan *et al.* reported other sulfated *N*-glycans bearing sulfated galactose, which were associated with the heavy chains of both human IgA subclasses but the specific *N*-glycosylation sites could not be determined yet [154]. These galactose-sulfated *N*-glycans were deposited in GlyConnect [186, 187]. Although the role of sulfated *N*-glycans in IgA remains largely unexplored, it is evident that they are important since they impact cell-cell interaction and, as HexNAc-sulfated sialosides, can be a ligand for some influenza hemagglutinin variants [293-297]. Despite their relevance, the site-specific detection of the low-abundant sulfated *N*-glycans by mass spectrometry is still pending due to their even lower abundance per glycosylation site, inadequate data analysis, and limitations associated with MS measurements (the negative charge of sulfated sugars causes low ionization efficiency and instability of the sulfated fragment ions) [185].

As demonstrated in chapter 4, the approach developed in the scope of this thesis allows the identification of intact *N*-glycopeptides from the low-abundant *N*-glycoproteome, including *N*-glycopeptides that feature sulfated, glucuronidated, or *O*-acetylated *N*-glycans [118, 226]. Therefore, the site-specific position of sulfated and other rare *N*-glycans in IgA was investigated by means of the in-depth *N*-glycoproteomic

workflow, as a first approach on the fractionated IgA chains and as a second approach on two commercially available human serum IgA samples (without protein fractionation) as shown in Figure 3.3. Multiple strategies to maximize rare *N*-glycopeptide identification were integrated in the data analysis, as presented in Figure 3.4 (section 3.2.6.3, Materials and Methods). For the first time, the site-specific identification of sulfated and *O*-acetylated *N*-glycans in human serum IgA was obtained.

This chapter presents that HexNAc-sulfated *N*-glycans are linked to the C_H2 domain sites N205, N131 in IgA2 and N144 in IgA1, as well as in the tailpiece site N340 in IgA1 and N327 in IgA2_{m1/n}. In addition to the sulfated *N*-glycan FA2G2S2-SO₄ (HexNAc₄Hex₅Fuc₁NeuAc₂Sulfo₁) described by the previous *N*-glycomics analyses conducted in this research group [44, 45], also other *N*-glycan compositions carrying sulfated HexNAc and *O*-acetylated sialic acid were identified (described in sections 5.2.1. and 5.2.2.) [44, 45]. *O*-acetylated *N*-glycans have been also characterized before in our group by Cajic *et al.* [45], but in horse serum. The application of the in-depth *N*-glycoproteomic workflow enabled the optimization of software-parameters for the identification of sulfated and *O*-acetylated *N*-glycopeptides in IgA. Later, the reproducible identification of these rare *N*-glycans was integrate into an expanded description *N*-glycosylation micro-heterogeneity of IgA. It was observed that the IgA tailpiece site shows the highest abundance of sulfated *N*-glycans. While *O*-acetylated *N*-glycans are present at a very low abundance. Also, the relative abundance of sulfated and *O*-acetylated *N*-glycans substantially differs between the commercial IgA samples. In the future, the methodology developed here will allow site-specific examination of sulfated *N*-glycans in IgA extracted from other body fluids (e.g., saliva or milk) [292]. Longitudinal studies that consider the analysis of sulfated *N*-glycans of IgA could be beneficial to improve the understanding of various clinical conditions such as inflammatory bowel diseases, rheumatoid arthritis, or nephropathy [183, 298-300]. Furthermore, expanding the research of the role of sulfated and *O*-acetylated *N*-glycans in IgA effector function can determine whether these rare *N*-glycans could be relevant as a critical quality attribute (CQA) of recombinantly expressed IgA for therapeutic use.

5.2. Results

Site-specific *N*-glycans on the Fc region from each human IgA subclass play a particular role in the effector function. This work focuses on the reliable site-specific identification of HexNAc-sulfated *N*-glycans on human serum IgA, primarily the FA2G2S2-SO₄ *N*-glycan, reported via glycomics analysis [44, 45]. As a first approach, the IgA heavy and light chains were separated (Figure 3.3), and their glycopeptide-enriched fractions were analyzed via MS/MS with the aim of identifying the *N*-glycosylation site(s) harboring the reported FA2G2S2-SO₄ *N*-glycan. An oxonium-ion-guided strategy was used to detect MS² spectra derived from sulfated *N*-glycopeptides, and focused software-assisted searches were conducted iteratively. This

first approach elucidated four sulfated *N*-glycan compositions, including the FA2G2S2-SO₄ *N*-glycan, and different search parameters for peptide identification. As a second approach, the glycoproteomic data analysis of human IgA was further expanded using two commercially available human serum IgA samples (without fractionation, Figure 3.4). Additional search strategies were applied to maximize the identification of rare *N*-glycans using not only oxonium marker ions of sulfated, and *O*-acetylated *N*-glycans (Figure 3.4B), but also wildcard searches and screening for other rare *N*-glycans (e.g., glucuronidated). *De novo* *N*-glycopeptide sequencing was conducted to report on newly discovered *N*-glycan compositions using the corresponding MS² spectra. In total, this work generated a reliable site-specific identification of thirteen rare *N*-glycans modified not only with sulfation but also with *O*-acetylation or glucuronidation, the last only in contaminant proteins. As a result, a comprehensive comparison of micro-heterogeneity of the *N*-glycosylation of these two human serum IgA samples was obtained. Figures 3.3 and 3.4 provide a visualization of the IgA derived samples and experimental strategies.

5.2.1. Site-specific identification of FA2G2S2-SO₄ *N*-glycan in the IgA heavy chain fraction

5.2.1.1. Screening for MS² spectra containing HexNAc-sulfated oxonium marker ions

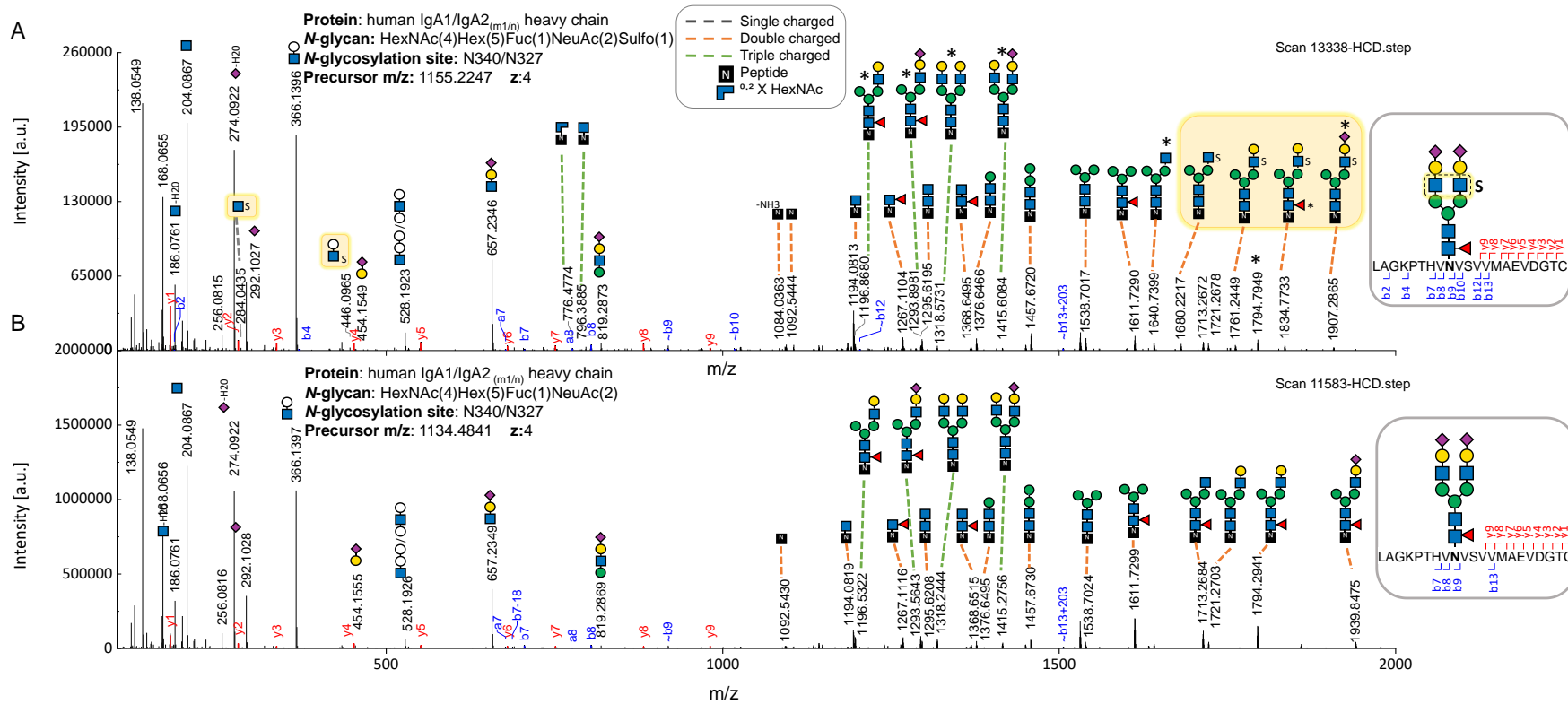
To reduce the complexity of human serum IgA, the light and heavy chains of human serum IgA were reduced, denatured, and fractionated as mentioned in chapter 3, section 3.2.2.2. (Supplementary Figure A.2.1 Appendix); both fractions were subsequently subjected to tryptic digestion and cotton-HILIC-SPE glycopeptide enrichment (section 3.2.4.2). The obtained glycopeptide enriched fractions were analyzed by MS and fragmented using stepped collisional energy. Both, precursor and fragment ions were measured by high-resolution mass spectrometry (section 3.2.6.). The data analysis first focused on identifying sulfated *N*-glycans in the MS² spectra. To this end, the MS² spectra containing HexNAc-sulfated oxonium marker ions (“sulfo-glycopeptide spectra”) were screened (section 3.2.6.2.). These screenings revealed 32 “sulfo-glycopeptide spectra” exclusively in the glycopeptide-enriched IgA fraction (Supplementary Table A.2.1 Appendix), apparently composed of IgA heavy chain subunits (Supplementary Figure A.2.1 Appendix). Sulfated *N*-glycopeptides were not found in the IgA fraction expected to contain the J- and light-chains of IgA1 and IgA2. This screening supports the assumption that the observed sulfated *N*-glycans are most likely attached to the heavy chains of the IgA subclasses [186, 187]. In addition to the sulfated *N*-glycan HexNAc₄Hex₅Fuc₁NeuAc₂Sulfo₁ (reported as FA2G2S2-SO₄ in previous glycomic studies conducted in our group [44, 45]), sulfated *N*-glycans with other compositions were identified using the GlycoMod tool (see section 3.2.6.2.) [233]. These new sulfated *N*-glycan compositions were integrated in subsequent *N*-glycoproteomic searches performed on IgA samples.

5.2.1.2. Glycoproteomic analysis focused on MS² spectra containing HexNAc-sulfated oxonium marker ions

One hypothesis for the source of sulfated *N*-glycans observed in glycomic analyses was that these *N*-glycans were attached to any of the *N*-glycosylation sites of the heavy chain constant region of serum IgA subclasses. Another hypothesis considers that sulfated *N*-glycans were attached to J-chain or plgR proteins –two proteins that form structural conformations with IgA in its dimeric form, which is reported to be present in low abundance in the blood serum (section 2.3.2. in chapter 2) [179]. In order to evaluate these hypotheses, the amino acid sequences of the IgA1 and IgA2 heavy chains, J-chain, and plgR were implemented in all glycoproteomic searches focusing on the identification of the “sulfo-glycopeptide spectra” acquired (Supplementary Table A.2.1 Appendix). This approach allowed the determination of the optimal search settings to match all previously detected “sulfo-glycopeptide spectra.” The resulting search parameter settings are shown in Supplementary Table A.2.2 (Appendix). The table also includes searches to identify *N*-glycopeptides with common *N*-glycans and non-glycopeptides. The resulting *N*-glycopeptide identifications were validated, as described in chapter 4 (Figure 4.3A) . The presence of HexNAc-sulfated oxonium marker ions was confirmed in the gPSMs containing sulfated *N*-glycans (displayed in the Supplementary Table A.2.4 Appendix) [226].

As summarized in Supplementary Table A.2.1 (Appendix), 29 “sulfo-glycopeptide spectra” could be assigned to 12 different sulfated *N*-glycopeptides with the tailpiece site (N340-IgA1/N327-IgA2_{m1/n}) bearing four different sulfated *N*-glycan compositions. These compositions include the one describing the FA2G2S2-SO₄ *N*-glycan plus three additional, which will be further described in section 5.2.5. Two more “sulfo-glycopeptide spectra” were assigned to one sulfated *N*-glycopeptide bearing the C_H2 domain site N205-IgA2_{m1/m2/n}. One “sulfo-glycopeptide spectrum” displayed a poor number of fragment ions, and it was not possible to identify the corresponding sulfated *N*-glycopeptide. Figure 5.2 shows b and y ion evidence of a peptide bearing a sulfated *N*-glycan at the tailpiece, compared to an identical glycopeptide without sulfation. It was also observed that most gPSMs with the tailpiece site N340-IgA1/N327-IgA2_{m1/n} harbor the sulfated *N*-glycan HexNAc₄Hex₅Fuc₁NeuAc₂Sulfo₁, which was structurally confirmed by *de novo* sequencing using HCD.step MS² spectra (Figure 5.2A). Comparing fragment spectra of both HexNAc-sulfated and non-sulfated *N*-glycan forms in Figure 5.2, the B and Y ions supporting the hypothesis of a structural difference (in the terminal HexNAc). Further evidence was obtained from *de novo* sequencing on a HCD.low scan acquired from the direct analysis of IgA and shown in Supplementary Figure A.2.2 (Appendix). This figure shows the oxonium ion HexNAc₁Hex₁NeuAc₁Sulfo₁ [M+H]⁺ confirming the position of the sulfated HexNAc at the antenna and not the core, which agrees with the structure of the sulfated *N*-glycan characterized by Cajic *et al.* and Chuzel *et al.* via glycomic analyses [44, 45].

Overall, by analyzing the IgA heavy chain fraction, the results demonstrate that the sulfated *N*-glycan FA2G2S2-SO₄ detected by glycomic analyses [44, 45], can be assigned to the IgA tailpiece homologous site and the N205-IgA2 in the C_H2 domain site. While other glycoproteomic studies have only associated galactose-sulfated *N*-glycans to the heavy chain region from IgA subclasses, yet without a site-specific description [154, 186, 187], the site-specific identification of sulfated *N*-glycans in IgA is shown here for the first time.



5.2.2. Expanded description of the *N*-glycosylation micro-heterogeneity of human serum IgA samples

As shown in Figure 3.4, human serum IgA samples from two commercial suppliers were directly analyzed (not fractionated by the GELFrEE® system) and strategies to maximize the identification of rare *N*-glycopeptides were applied to achieve an expanded *N*-glycoproteomic analysis. To this end, the tryptic-digested samples from both suppliers were enriched for glycopeptides. Technical quadruplicates were obtained for each commercial sample by repeating the cotton-HILIC-SPE protocol four times. All HILIC-fractions from the technical replicates (loading + wash, elution 1, and elution 2) were analyzed by nanoRP-LC-ESI-OT-OT-MS/MS (section 3.2.6.1.). All measurements of the HILIC-fractions were included in the subsequent proteomic and *N*-glycoproteomic searches.

5.2.2.1. Site-specific identification of HexNAc-sulfated *N*-glycans

In addition to the previously reported sulfated *N*-glycan, FA2G2S2-SO₄ [44, 45], six new *N*-glycan compositions presenting HexNAc-sulfated oxonium marker ions were identified at the C_H2 domain and tailpiece *N*-glycosylation sites of IgA. As described in chapter 3 (section 3.2.6.), the identification of the *N*-glycan compositions was achieved not only from the analysis of fractionated IgA but also from the direct analysis of human IgA samples in combination with the application of strategies to maximize the identification of unknown *N*-glycan compositions with sulfation. The GlyConnect *N*-glycan databases belonging to IgA1 and IgA2 (UniProtKB entries: P01876 and P01877, respectively) were considered as a source of sulfated *N*-glycan compositions (Table 5.1) [262, 264]. Next, a *N*-glycoproteomic analysis focused on including HexNAc-sulfated *N*-glycan compositions was conducted using settings listed in Supplementary Table A.2.3 (Appendix). The structures of the identified sulfated *N*-glycans were described to some extent, through HCD.low and HCD.step fragment ion spectra. The Supplementary Figures A.2.3 to A.2.8 (Appendix) display the manual annotation of six sulfated *N*-glycans, additional to the previously reported sulfated *N*-glycan, FA2G2S2-SO₄. These spectra revealed one HexNAc-sulfated di-fucosylated hybrid-type *N*-glycan, two HexNAc-sulfated sialylated complex-type *N*-glycans without core fucosylation, one HexNAc-sulfated sialylated bisected complex-type *N*-glycan with core fucosylation, and one HexNAc-sulfated mono-sialylated complex-type *N*-glycan with core fucosylation. A sixth sialylated and core fucosylated complex-type *N*-glycan holding two sulfated monosaccharides (di-sulfated) was identified (Supplementary Figure A.2.5 Appendix); while sulfated HexNAc is evident, it was not possible to confirm the position of the second sulfation. As Figure 5.3 shows, this work not only describes a larger list of sulfated *N*-glycans but also provides a site-specific overview of IgA sulfated *N*-glycans for the first time. In this Figure, it is observed that the seven sulfated *N*-glycan compositions were identified in both commercial IgA samples.

5. Site-specific identification of sulfated and *O*-acetylated *N*-glycans in human serum IgA

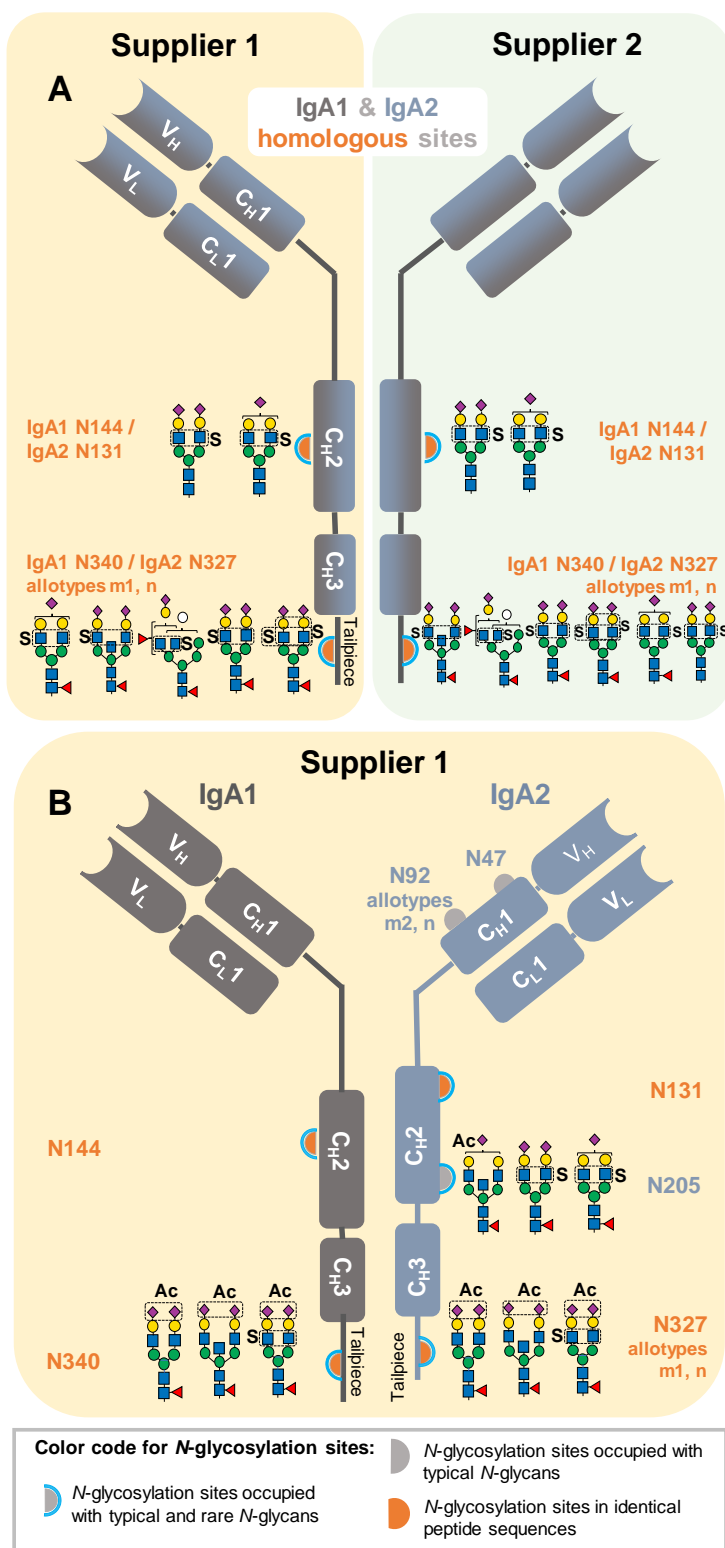
Table 5.1. Cross-referenced *N*-glycans bearing a sulfated-sugar and *O*-acetylated NeuAc.

| | <i>N</i> -glycan compositions | Observed in this study | Evidence reported in GlyConnect | | |
|----|--|------------------------|---------------------------------|------------------|------------|
| | | Evidence | Evidence | Protein | Reference |
| 1 | HexNAc(5)Hex(5)Fuc(1)NeuAc(2)Ac(1) | NeuAc acetylation | No | - | - |
| 2 | HexNAc(5)Hex(5)Fuc(1)NeuAc(1)Ac(1) | NeuAc acetylation | No | - | - |
| 3 | HexNAc(4)Hex(5)Fuc(1)NeuAc(2)Ac(1)* | NeuAc acetylation | NeuAc acetylation | IgA1 Hc | [186] |
| 4 | HexNAc(4)Hex(5)Fuc(1)NeuAc(2)Ac(1)Sulfo(1) | GlcNAc sulfation | No | - | - |
| 5 | HexNAc(4)Hex(5)Fuc(1)NeuAc(2)Sulfo(2) | GlcNAc sulfation | No | - | - |
| 6 | HexNAc(5)Hex(5)Fuc(1)NeuAc(2)Sulfo(1) | GlcNAc sulfation | No | - | - |
| 7 | HexNAc(4)Hex(5)Fuc(1)NeuAc(2)Sulfo(1)* | GlcNAc sulfation | Gal sulfation | IgA1 Hc, IgA2 Hc | [186, 187] |
| 8 | HexNAc(4)Hex(5)Fuc(1)NeuAc(1)Sulfo(1)* | GlcNAc sulfation | Gal sulfation | IgA1 Hc | [186] |
| 9 | HexNAc(4)Hex(5)NeuAc(2)Sulfo(1)* | GlcNAc sulfation | Gal sulfation | IgA2 Hc | [187] |
| 10 | HexNAc(4)Hex(5)NeuAc(1)Sulfo(1) | GlcNAc sulfation | No | - | - |
| 11 | HexNAc(4)Hex(6)Fuc(2)NeuAc(1)Sulfo(1) | GlcNAc sulfation | No | - | - |
| 12 | HexNAc(4)Hex(5)Sulfo(1) | Not observed | GlcNAc sulfation | IgA1 Hc | [186] |

* *N*-glycan compositions found in this study and also reported in GlyConnect [154, 186, 187]. Hc: Heavy chain.

5.2.2.2. Site-specific identification of *O*-acetylated *N*-glycans

As shown in Table 5.1, one *N*-glycan bearing *O*-acetylated sialic acid and annotated as HexNAc₄Hex₅Fuc₁NeuAc₂Ac₁, has been reported to IgA1 before [154, 186, 187]. Therefore, MS² scans were screened for NeuAc *O*-acetylated oxonium marker ions. Once their presence was confirmed, the strategies to maximize the identification of unknown *N*-glycan compositions with *O*-acetylated NeuAc, described in section 3.2.6.3. were applied. Then, a specific *N*-glycoproteomic search, including *N*-glycan compositions with *O*-acetylated NeuAc, was conducted using the parameters listed in Supplementary Table A.2.3 (Appendix). This resulted in the identification of the reported *O*-acetylated *N*-glycan at the site N340-IgA1/N327-IgA2_{m1/n} (Supplementary Figure A.2.9 Appendix), plus three new *N*-glycan compositions bearing *O*-acetylated NeuAc. The three additional NeuAc *O*-acetylated *N*-glycans detected were *de novo* sequenced to describe them in detail (Supplementary Figures A.2.10 to A.2.12 Appendix).



As displayed in Figure 5.3, one *O*-acetylated mono-sialylated *N*-glycan composition bearing core fucosylation and bisecting HexNAc was identified at the unique IgA2 C_H2 domain site N205. Then, three complex-type *O*-acetylated *N*-glycans bearing core fucosylation were identified at the tailpiece homologous site. Interestingly, all of the NeuAc *O*-acetylated *N*-glycans were only identified in the human serum IgA sample from supplier 1. Surprisingly, one of the *O*-acetylated *N*-glycans at the tailpiece site bears also sulfated HexNAc. This very rare *N*-glycan was further interpreted by manual *de novo* sequencing in the HCD.low MS² spectrum (Supplementary Figure A.2.12A Appendix). Its HCD.low MS² spectrum not only presents oxonium ions of *O*-acetylated sialic acid; but also HexNAc₁Sulfo₁ [M+H]⁺ oxonium ion and Y ions demonstrating sulfation at the terminal HexNAc. Moreover, the HexNAc₁Sulfo₁ [M+H]⁺ oxonium ion was also evident in the HCD.step MS² spectrum (Supplementary Figure A.2.12B Appendix).

Figure 5.3. Site-specific description of the here identified *N*-glycans carrying sulfated HexNAc and *O*-acetylated NeuAc in human serum IgA. (A) Sulfated *N*-glycans identified in IgA1 and IgA2 homologous sites (B) Sulfated *N*-glycans identified in unique *N*-glycosylation sites of IgA2 and NeuAc *O*-acetylated *N*-glycans identified within the entire analysis. The sample where each *N*-glycan composition was identified is also indicated for IgA sample from supplier 1 and 2 in yellow and green, respectively.

5.2.2.3. Peptide modifications detected in IgA *N*-glycopeptides

A key adjustment for the improvement of the *N*-glycopeptide search was the identification of peptide modifications. Multiple gPSMs with the tailpiece *N*-glycosylation site N340-IgA1/N327-IgA2_{m1/n} correspond to different mass variations of the predicted tryptic peptide (Supplementary Tables A.2.4 and A.2.5 Appendix). It was found that these shifts in peptide mass were caused by up to two amino acid oxidations, one missed cleavage, or truncated tailpiece forms. Methionine was the most frequently oxidized amino acid, and even dioxidation was observed in later analyses. The most frequent truncated tailpiece form lacks the C-terminal tyrosine—this tyrosine truncated tailpiece form has been previously reported [7, 290, 301]. Interestingly, several gPSM containing the tailpiece site were identified via a semi-specific tryptic search, matching a C-terminal ragged peptides. One example is the peptide LAGKPTHVNVS.V lacking the last 11 amino acids of the tailpiece (VVMAEVDGTCY). Studies have demonstrated that the penultimate cysteine (Cys352) is critical for IgA dimerization and disulfide bond formation with J-chain and other proteins [302-304]. Clarifying whether this short peptide results from an unspecific cleavage or is another truncated tailpiece variant of the IgA subclasses requires further investigation.

5.2.2.5. Identification of rare *N*-glycopeptides from “contaminant” proteins

The screening for MS² scans containing rare oxonium ions was conducted using the list of oxonium marker ions reported in chapter 4 (Table 4.2). The oxonium ion Hex₁HexNAc₁HexA₁ [M+H]⁺, (*m/z* 542.1716) was detected in the serum IgA samples from both suppliers. The experimental *m/z* of the oxonium ion observed closely matches the theoretical *m/z* of Hex₁HexNAc₁HexA₁ ion, supporting the hypothesis that a HexA is present as capping sugar. This oxonium ion serves as an indicator of *N*-glycans carrying HNK-1 glycoepitope, which is present in human brain *N*-glycans and bears sulfated glucuronic acid [251]. Therefore, it is assumed that the type of hexuronic acid observed in our study is glucuronic acid [29, 251]. Following the strategies for maximizing the identification of unknown *N*-glycan compositions (described in chapter 3, section 3.2.6.3.), two glucuronidated *N*-glycan compositions were identified, one of which also contained sulfation as a modification. Then, the *N*-glycoproteomic search revealed that these glucuronidated *N*-glycans do not belong to any *N*-glycosylation site of IgA1 or IgA2 but to α-1-antitrypsin (N271) and α-2-HS glycoprotein (N156) (Supplementary Table A.2.5 Appendix). To elucidate their *N*-glycan structure, the *N*-glycopeptides were *de novo* sequenced (Supplementary Figures A.2.13 and A.2.14 Appendix). Although both *N*-glycopeptides carry di-antennary complex-type *N*-glycans with one sialic acid and one glucuronic acid as capping sugars, the α-2-HS-glycoprotein *N*-glycopeptide showed the oxonium ion Hex₁HexNAc₁HexA₁Sulfo₁ [M+H]⁺, thereby confirming the presence of the sulfated glycoepitope HNK-1.

HexNAc-sulfated oxonium marker ions were also detected in gPSMs of other proteins (Supplementary

Table A.2.5 Appendix), such as Immunoglobulin J-chain bearing HexNAc₄Hex₅NeuAc₂Sulfo₁ *N*-glycan at N71 site. Also, the sulfated *N*-glycan HexNAc₄Hex₅Fuc₁NeuAc₂Sulfo₁, previously reported via glycomics, was identified on *N*-glycopeptides from α -2-HS-glycoprotein (N156) and complement C3 (N85), however their gPSMs present a low number of b and y ions.

5.2.2.6. Evaluation of the relative abundance of each IgA subclass

As described in chapter 3 (section 3.2.6.3.), the relative abundance of each IgA subclass was calculated based on the quantification of selected peptides identified by Proteome Discoverer™. Supplementary Table A.2.6 (Appendix) provides the peptide identification list. Around 80% of the protein sequence of IgA1 and IgA2 is homologous. Therefore, some peptides are assigned to both IgA subclasses, which can lead to misleading results. To avoid bias caused by tryptic peptides with homologous IgA sequences (ambiguous peptides), the non-glycopeptides “NFPPSQDASGDLYTTSSQLTLPATQCLAGK” and “NFPPSQDASGDLYTTSSQLTLPATQCPDGK,” were selected to represent IgA1 and IgA2, respectively. It was found that the ratio between IgA1:IgA2 subclasses was 89:11 in the commercial sample from supplier 1, which is the natural ratio in blood serum [178], and approximately 99:1 in the one from supplier 2, which is ten times higher (Table 5.2). This abnormal ratio in the sample from supplier 2 suggests an enrichment of IgA1, possibly as a consequence of the IgA purification process employed (e.g., immobilized jacalin chromatography) [305].

Table 5.2. Relative abundance of IgA subclasses of each human serum IgA commercial sample.

| Supplier | IgA subclass | UniProtKB accession number | Replicates | | | | Relative abundance* | |
|----------|--------------|----------------------------|------------|-------|-------|-------|---------------------|------|
| | | | a | b | c | d | Mean | SD |
| 1 | IgA1 | P01876 | 86% | 89% | 90% | 91% | 89.0% | 2.1% |
| | IgA2 | P01877, P0DOX2 | 14% | 11% | 10% | 9% | 11.0% | 2.1% |
| 2 | IgA1 | P01876 | 99.2% | 98.6% | 99.6% | 99.4% | 99.2% | 0.4% |
| | IgA2 | P01877, P0DOX2 | 0.8% | 1.4% | 0.4% | 0.6% | 0.8% | 0.4% |

*AUC XIC of the reference peptide of an IgA subclass/total AUC XIC of the reference peptides from both IgA subclasses identified in each commercial sample.

5.2.2.7. Expanded *N*-glycoproteomic analysis of IgA: integration of common and rare *N*-glycans

In order to maximize the site-specific description of IgA *N*-glycosylation, an expanded *N*-glycoproteomic analysis was performed integrating the new insights on sulfated, *O*-acetylated, and glucuronidated *N*-glycan compositions, as well as peptide modifications. Therefore, a proteomic search and multiple *N*-glycoproteomic searches were conducted featuring rare and common *N*-glycan compositions using Byonic™ search engine and implementing oxonium ions MS/MS filters specific to each *N*-glycan

modification (Supplementary Table A.2.3 Appendix). To include (*N*-glyco-)peptides from “contaminant” proteins, the human canonical proteome UniProtKB was set as protein database for the proteomic search, and a focused database was extracted from each sample based on the proteins identified. (*N*-glyco-)peptide identifications with a “Byonic™ MS2 search score” below 100 were excluded due to their poor quality, and the remaining *N*-glycopeptide identifications were manually validated, as described in Figure 4.3. During validation, gPSM presenting rare *N*-glycans were accepted as “True” if their corresponding oxonium marker ions were present in the MS² spectra. Supplementary Table A.2.5 (Appendix) shows peptides and *N*-glycopeptides identified within all replicates from both commercial samples, which derive from IgA subclasses but also “contaminant” proteins. Most of the “contaminant” *N*-glycopeptides belong to α -1-antitrypsin, complement C3, kininogen-1, and α -2-HS-glycoprotein.

Figure 5.4 displays a comparative relative quantification of the *N*-glycosylation micro-heterogeneity of the human serum IgA samples from both suppliers. Considering that identical tryptic peptides represent the IgA tailpiece site (N340-IgA1/N327-IgA2_{m1/n}) and the C_H2 domain site (N144-IgA1/N131-IgA2), the micro-heterogeneity analysis on those *N*-glycosylation sites is assigned to both IgA subclasses (Supplementary Tables A.2.7 and A.2.8 Appendix). The micro-heterogeneity analysis for unique *N*-glycosylation sites of IgA2 is listed separately for each sample in Supplementary Tables A.2.9 and A.2.10 (Appendix).

As a first step, it was essential to confirm that this expanded *N*-glycoproteomic analysis reflects the typical high abundant *N*-glycans reported for each *N*-glycosylation site of the human serum IgA subclasses [7, 110, 152, 183]. The analysis shows that while the tailpiece site (N340-IgA1/N327-IgA2_{m1/n}) harbors core fucosylated sialylated complex-type di- and multi-antennary *N*-glycans, the C_H2 domain site (N144-IgA1/N131-IgA2) bears mostly non-fucosylated sialylated hybrid- and di-antennary complex-type *N*-glycans (Figure 5.4A). Thus, *N*-glycans observed in other glycoproteomic analyses that included the simultaneous analysis of both IgA subclasses [7, 110, 152, 183], were also observed within this expanded *N*-glycoproteomic analysis at the corresponding sites in the C_H2 domain and in the tailpiece. Also, in agreement with a preceding study, Figure 5.4B shows that *N*-glycosylation sites unique from the IgA2 subclass (N92-IgA2_{m2/n} and N205-IgA2_{m1/m2/n}) predominantly display sialylated core-fucosylated complex-type *N*-glycans, including bisecting *N*-glycans [152]. This demonstrates the consistency of this analysis with other works. In regard to other *N*-glycosylation sites unique from the IgA2 subclass, one *N*-glycopeptide representing the N47-IgA2 site with a mono-sialylated core-fucosylated complex-type *N*-glycan and a few *N*-glycopeptides presenting complex-type *N*-glycans at the N327-IgA2_{m2} were detected.

As a next step, the newly identified rare *N*-glycan compositions were evaluated. Three of the four *O*-acetylated *N*-glycan compositions detected returned a peak area in more than three replicates of IgA sample from supplier 1, and their relative abundance was estimated here for the first time. The expanded

glycoproteomic analysis reflects that NeuAc *O*-acetylated complex-type *N*-glycans were present at a very low relative abundance (Figure 5.4). At the tailpiece site N340-IgA1/N327-IgA2_{m1/n}, the NeuAc *O*-acetylated bisected complex-type *N*-glycan and the non-bisected complex-type *N*-glycan display a total relative abundance of 0.09% (Supplementary Table A.2.7 Appendix). At the C_H2 domain site N205-IgA2 (Figure 5.4B), a mono-sialylated bisected NeuAc *O*-acetylated complex-type *N*-glycan was observed with an average relative abundance of 0.03% (Supplementary Table A.2.9 Appendix). The *N*-glycan bearing both NeuAc *O*-acetylated and sulfated HexNAc identified at the tailpiece site appeared only in one technical replicate of the sample from supplier 1, therefore, it was not quantified.

Figure 5.4A also shows that the sulfated *N*-glycan (previously reported by our group via glycomic analyses FA2G2S2-SO₄ [44, 45]) was identified with the highest abundance at the tailpiece *N*-glycosylation site N340-IgA1/N327-IgA2_{m1/n} (0.74% and 3.85% in the sample from supplier 1 and supplier 2, respectively). In contrast, the same *N*-glycan was detected in a very low abundance at the C_H2 domain site N205-IgA2 (only 0.03% of the total *N*-glycans in the sample from supplier 1, Supplementary Table A.2.9 Appendix). At the homologous site in C_H2, N144-IgA1/N131-IgA2, the samples from both suppliers presented sulfated non-fucosylated complex-type *N*-glycans in a very low relative abundance (Supplementary Tables A.2.7 and A.2.8 Appendix). In fact, only the sample from supplier 2 returned an area under the curve value for at least three technical replicates, summing 0.05% of sulfated *N*-glycans at the C_H2 domain homologous site (Supplementary Table A.2.8 Appendix). The di-sulfated complex-type *N*-glycan HexNAc₄Hex₅Fuc₁NeuAc₂Sulfo₂, detected at the tailpiece site, shows 0.01% and 0.54% in the sample from supplier 1 and supplier 2, respectively. Overall, while the IgA sample from supplier 2 shows the highest relative abundance of sulfated *N*-glycans at the IgA1/IgA2 homologous sites, only the sample from supplier 1 presents peak area values of sulfated *N*-glycans that belong only to IgA2 *N*-glycosylation sites.

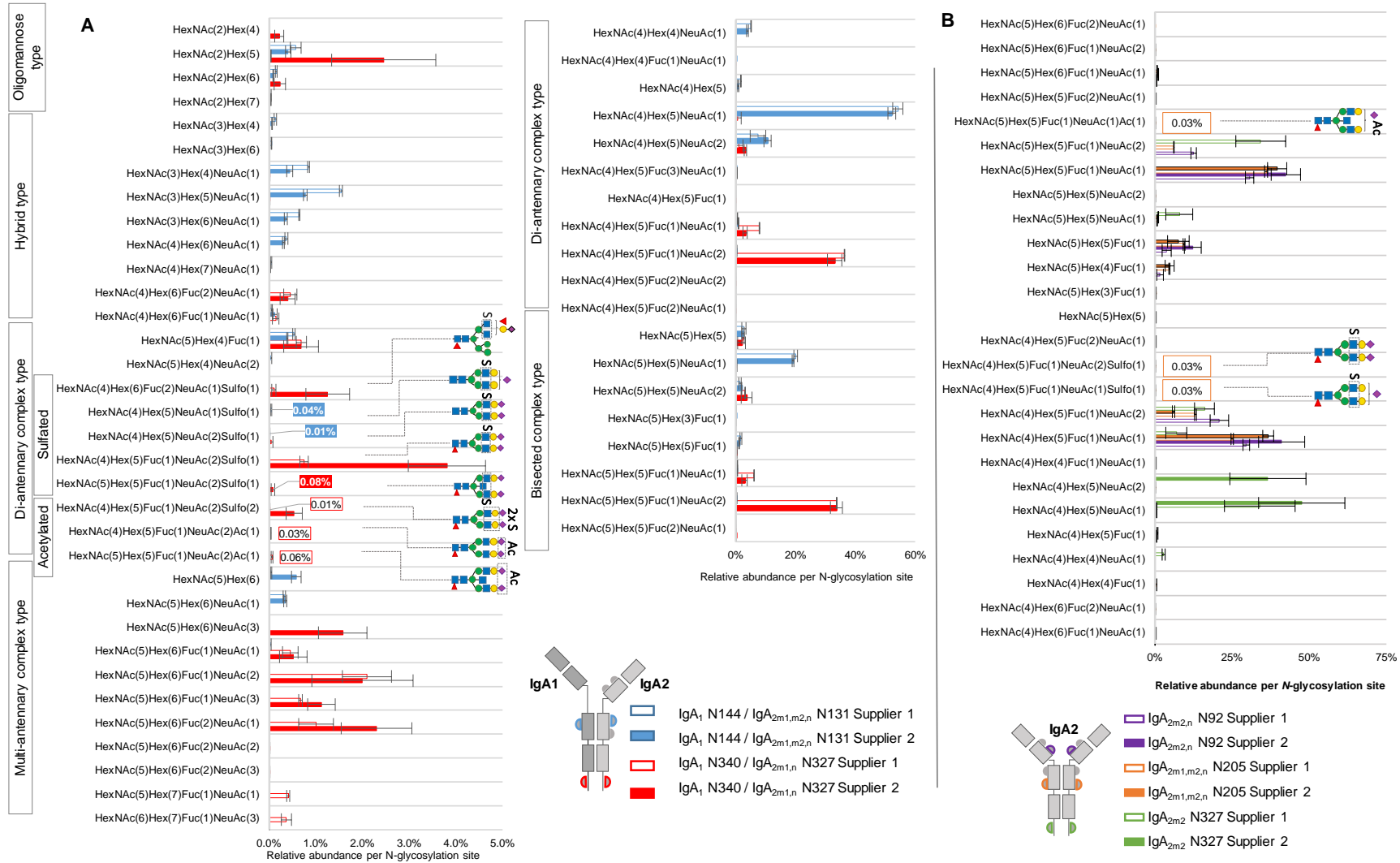


Figure 5.4. Comparison of micro-heterogeneity of N-glycosylation identified in commercially available human serum IgA samples from two suppliers. A) Relative abundance of N-glycans at the homologous N-glycosylation sites N340-IgA1/N327-IgA2_(m1/n) and N144-IgA1 / N131-IgA2. B) Relative abundance of N-glycans at the N-glycosylation sites belonging only to the IgA2 subclasses. Results for relative quantification are shown in the Supplementary Tables A.2.7-A.2.10 (Appendix).

5.2.2.8. Comparison of two data analysis approaches for the identification of sulfated *N*-glycopeptides

To address the challenge of identifying oxonium fragment ions derived from sulfated *N*-glycans, a second set of *N*-glycopeptide searches without applying sulfated oxonium ions as part of the MS/MS filter was conducted and manually validated. The sulfated *N*-glycopeptides identified included sulfated gPSM that exhibited the presence or absence of HexNAc-sulfated oxonium ions in the MS² spectra (Supplementary Table A.2.11 Appendix). During manual validation, the sulfated gPSM lacking HexNAc-sulfated oxonium ions were assigned as “True” if the match was correct regarding the b and y fragment ions, the conserved peptide fragmentation pattern, and common glycan oxonium ions (e.g. HexNAc₁ [M+H]⁺, NeuAc₁ [M+H]⁺, HexNAc₁Hex₁ [M+H]⁺). Then, gPSM bearing sulfated *N*-glycans were classified by the presence or absence of HexNAc-sulfated oxonium marker ions. As the homologous tailpiece site (N340-IgA1/N327-IgA2_{m1/n}) is the primary source of sulfated *N*-glycans, the relative quantification of sulfated *N*-glycans focused exclusively on this site (Supplementary Table A.2.12 Appendix). The abundance of each *N*-glycan composition was normalized by the total area under the curve of all *N*-glycan compositions found at the homologous tailpiece site and it is displayed in Figure 5.5.

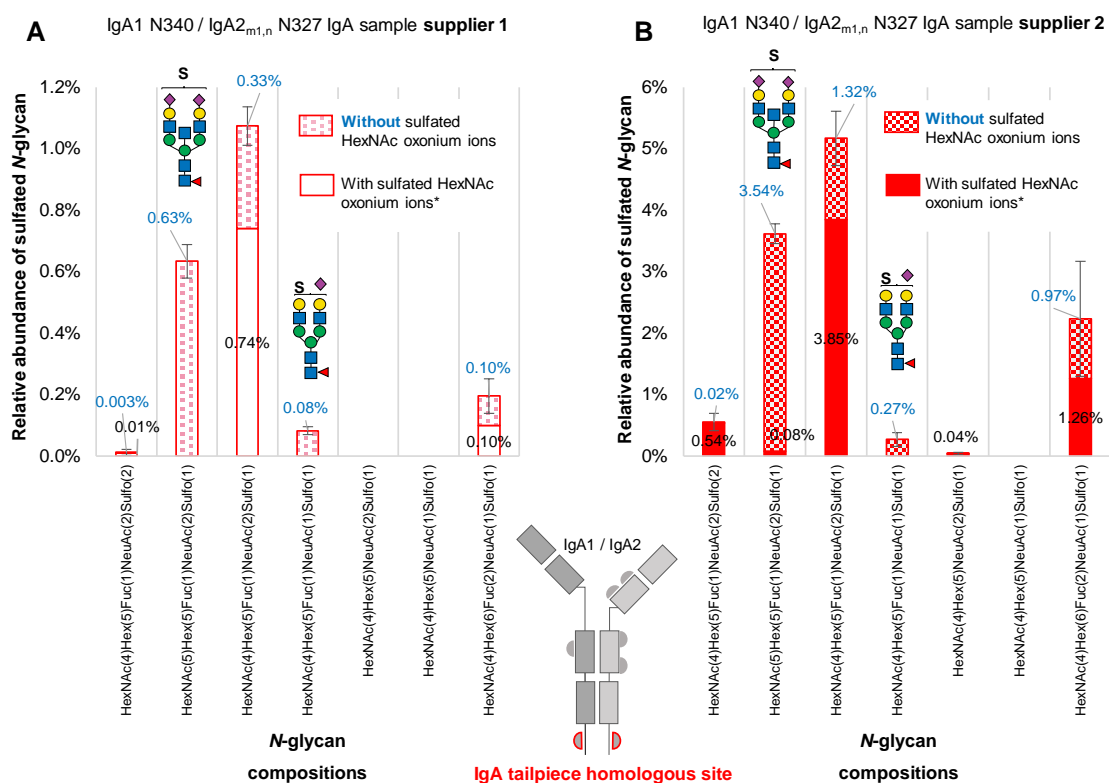


Figure 5.5. A comparative analysis of sulfated *N*-glycans with and without HexNAc-sulfated oxonium ions attached to the IgA tailpiece site in commercially available human serum IgA samples from two suppliers. **(A)** The sample from supplier 1. **(B)** The sample from supplier 2. Sulfated *N*-glycopeptides were identified, manually validated, and verified for oxonium marker ions through Byonic™. Two sulfated *N*-glycan structures, one with bisecting HexNAc and another mono-sialylated, appear more

abundant in glycopeptide spectra matches when their spectra do not present oxonium marker ions of sulfated HexNAc. Asterisk () indicates that oxonium-ion-based evidence refers to detection of $\text{Sulfo}_1\text{HexNAc}_1[\text{M}+\text{H}]^+$ or $\text{Sulfo}_1\text{HexNAc}_1\text{Hex}_1[\text{M}+\text{H}]^+$ oxonium ions.*

As can be observed in Figure 5.5, there is an increase of approximately 30% in the relative abundance of the composition $\text{HexNAc}_4\text{Hex}_5\text{Fuc}_1\text{NeuAc}_2\text{Sulfo}_1$ (related to the reported *N*-glycan FA2G2S2-SO₄ [44, 45]) in both samples. One of the biggest differences between the sulfated *N*-glycopeptide identified within each glycoproteomic search is the presence of the sulfated bisected *N*-glycan $\text{HexNAc}_5\text{Hex}_5\text{Fuc}_1\text{NeuAc}_2\text{Sulfo}_1$. With the relative quantification of sulfated *N*-glycopeptides presenting HexNAc oxonium ions, the sulfated bisected *N*-glycan only appears in one of the replicates of the sample from supplier 1, and with very low relative abundance (0.08%) in the sample from supplier 2. However, by considering sulfated *N*-glycopeptides lacking HexNAc oxonium ions, this bisected *N*-glycan becomes the second most predominant sulfated *N*-glycan in both samples, reflecting an increase of 0.63% and 3.54% in the samples from supplier 1 and supplier 2, respectively. Although the *N*-glycopeptide identifications bearing the sulfated bisected *N*-glycan lacking HexNAc oxonium ions are detected within independent chromatographic peaks, the average scan time of the precursor ions representing those peaks is not significantly different from the identifications presenting HexNAc-sulfated oxonium ions. The case is similar for the sulfated mono-sialylated *N*-glycan, whose relative abundance becomes evident by including sulfated *N*-glycopeptides lacking HexNAc oxonium ions.

5.3. Discussion

This investigation is the first to include in total seven HexNAc-sulfated plus four NeuAc *O*-acetylated *N*-glycan compositions in the site-specific description and relative quantification of IgA *N*-glycosylation. Overall, both commercial samples consistently showed that the primary source of HexNAc-sulfated *N*-glycans is the *N*-glycosylation site at the IgA tailpiece (N340-IgA1/N327-IgA2_{m1/n}). GlyConnect *N*-glycan databases from IgA1 and IgA2 showed only one of the four NeuAc *O*-acetylated *N*-glycans and three of the seven sulfated *N*-glycans detected here. However, these three *N*-glycans were reported to be modified with galactose sulfation instead of GlcNAc sulfation [154, 186, 187]. The glycoproteomic analysis conducted hereby does not provide any probative structural information on galactose sulfation for two reasons. First, the oxonium marker ion $\text{Hex}_1\text{Sulfo}_1[\text{M}+\text{H}]^+$, potentially generated by a sulfated galactose, was never detected in any of the gPSMs containing sulfated *N*-glycans. Second, as She et al. reported, detection of galactose sulfation in intact *N*-glycopeptides relies on the manual identification of Y ions holding terminal galactose sulfation [245]. Although manual validation can pursue the identification of galactose-sulfated *N*-glycans, software tools capable of comprehensive Y ions matching, can facilitate this identification process. Therefore, this type of identification requires establishing a reliable and less time-

consuming strategy for identifying the expected Y ions [245]. For example, the GlycanFinder software performs de novo sequencing of *N*-glycopeptides and can be customized to identify peptides bearing galactose-sulfated *N*-glycans [306].

The *N*-glycoproteomic analysis conducted without using sulfated oxonium ions as part of the MS/MS filter revealed correct gPSM containing sulfated *N*-glycan compositions but lacking HexNAc-sulfated oxonium marker ions. By considering this type of gPSM, the bisected sulfated *N*-glycan turned the second most abundant *N*-glycan at the tailpiece site. This outcome points to a broader picture. On the one hand, it is possible that these differences in sulfated *N*-glycopeptide abundances are caused by the instability of the HexNAc-sulfated oxonium ions after fragmentation, generating MS² spectra lacking these ions [185]. On the other hand, it is possible that different sulfation modifications coexist in the IgA sample, like galactose sulfated *N*-glycans, which has also been reported in human IgA via glycomic analyses [154, 186, 187].

It was observed that the relative abundance of NeuAc *O*-acetylated *N*-glycans at the tailpiece *N*-glycosylation site is at least 10 times lower than the abundance of sulfated *N*-glycans in the IgA tailpiece for sample from supplier 1. Surprisingly, no *O*-acetylated *N*-glycans were detected in the sample from supplier 2, which showed a massive depletion of IgA2 subclass—present within an IgA1:IgA2 ratio 10 times lower than the normal ratio in human serum [178]. This ratio can be affected by the purification process applied [305]. Differences in *N*-glycosylation between both IgA subclasses have been demonstrated in preceding studies. By separation of IgA1 and IgA2 subclasses, Steffen et al. described that the total IgA2 *N*-glycome shows low levels of sialylated bisected *N*-glycans and higher levels of hybrid- and oligomannose-type *N*-glycans [184]. Chandler et al. also showed by the specific glycoproteomic analysis of IgA2, that the C_H2 domain site N131-IgA2 bears hybrid-type *N*-glycans in higher abundance than the C_H2 domain site N144 in IgA1 [152]. These studies suggest that the share of hybrid-type *N*-glycans will be influenced by the IgA1:IgA2 ratio, which might explain the differences observed in our analysis, between both commercial samples. A question that also arises is to what extent the IgA1:IgA2 subclass ratio affects the abundance of *O*-acetylated and sulfated *N*-glycans. If the substantial difference in abundance of sulfated *N*-glycans observed between IgA samples is related to the IgA1:IgA2 ratio present in each sample, this would suggest that sulfated *N*-glycans are selectively conserved at the tailpiece *N*-glycosylation site in IgA1 and not in IgA2. Nonetheless, further analyses are required to elucidate differences in the dominance of sulfated *N*-glycans between the IgA subclasses.

Rare *N*-glycans bearing sulfated HexNAc or HNK-1 glycoepitope were detected in a few contaminant proteins. The presence of these “contaminant” proteins in the purified IgA is probably due to chromatographic co-elution or to stable protein-protein interactions during the purification process [305, 307, 308]. Studies showed that proteins like complement C3 and β -2-glycoprotein form circulating immune

complexes with IgA subclasses in blood serum [307, 308]. Additionally, a strong interaction is controlled by the penultimate cysteine of the IgA tailpiece, forming a disulfide bond either with albumin or α -1-antitrypsin [304]. Therefore, it is essential to consider the contribution of *N*-glycans from “contaminant” proteins in glycomic analysis.

Little is known about the role of NeuAc *O*-acetylated and sulfated glycans in the effector function of each IgA subclass. On the one hand, sulfated glycoepitopes are ligands for L-selectin, and this interaction plays a vital role in cell adhesion and trafficking of immune system cells [89, 296, 297, 309]. On the other hand, studies that used glycan arrays demonstrated the affinity of specific influenza viruses (and also bacteria) for sialylated glycoepitopes containing sulfated GlcNAc [293, 310, 311]. Also, NeuAc *O*-acetylation influences the binding to viral neuraminidases during infection [312], and may have immunoregulatory effects by affecting the affinity of the sialylated glycan for lectins (e.g., CD22) or bacterial sialidases [313, 314]. It has been reported that sialylated *N*-glycans at the IgA tailpiece play a role in neutralizing influenza virus infection [172]. Thus, it is possible that IgA sulfated and *O*-acetylated *N*-glycans also play a crucial role in the IgA antiviral activity, for instance, by mimicking a cell receptor-ligand used during infection.

New methods have been developed for the detailed glycomic analysis of sulfated and *O*-acetylated *N*-glycans. As mentioned in section 5.1., Cajic *et al.* established and applied a technique for multiple analyses of *N*-glycans employing a removable fluorescent labeling (Fmoc) [45]. Using this glycomic workflow, they captured and exhaustively characterized *O*-acetylated NeuAc *N*-glycans (from horse serum proteins) and the sulfated *N*-glycan FA2G2S2-SO₄ from human serum IgA. In addition, the glycomic workflow established by Cajic *et al.* was combined with a highly specific sulfatase and a sulfate-dependent hexosaminidase, characterized by Chuzel *et al.* [44]. Later Chuzel *et al.* also confirmed that this sulfatase can act as a highly selective lectin for GlcNAc-6-SO₄ in the absence of calcium. Although the main limitation of this glycomic workflow is processing time, the *N*-glycan structure is highly elucidated. The glycoproteomic workflow presented in this study has, in contrast, limitations regarding structure elucidation. Nevertheless, the site-specific description of rare *N*-glycan compositions, can already improve the exploration of IgA *N*-glycosylation in biological or clinical studies. Considering that differences in IgA *N*-glycosylation analyzed from various body fluids (e.g., serum, saliva and human milk) have been reported [110, 292], it would be useful to apply our glycoproteomic workflow to IgA extracted from other body fluids. Overall, more investigations are required to clarify the function of sulfated and *O*-acetylated *N*-glycans in the protein-glycan interaction network and to evaluate if these are relevant, e.g., for IgA production as a therapeutic or clinical practice as a diagnostic tool.

5.4. Summary

Sulfated *N*-glycans of human immunoglobulin A (IgA) were recently discovered via glycomic approaches. However, a site-specific description was pending. Certain *N*-glycan compositions at specific *N*-glycosylation sites in IgA are crucial for microbial neutralization and effector functions. For instance, sialylated *N*-glycans of the C-terminal tailpiece mediate anti-viral activity by interfering with sialic-acid-binding viruses. Sulfated *N*-glycan epitopes can be ligands for viral proteins and thus play a role in the immune response. In this study, a site-specific screening for sulfated and other rare *N*-glycans in two commercially available human serum IgA samples was performed, employing an in-depth *N*-glycoproteomic approach previously developed here. Evidence of complex-type and hybrid-type *N*-glycans carrying sulfated HexNAc attached to the *N*-glycosylation sites in the tailpiece and the C_H2 domain of both IgA subclasses was found. A detailed comparison of the *N*-glycosylation profiles of human serum IgA samples from two suppliers showed HexNAc-sulfated *N*-glycans consistently in high abundance (at the *N*-glycosylation site) in the tailpiece region. Also, complex-type *N*-glycan compositions bearing *O*-acetylated sialic acid were identified at the tailpiece and the C_H2 domain sites. Surprisingly, *N*-glycans bearing glucuronic acid were identified in both commercial IgA samples, but from peptides of “contaminant” glycoproteins. These findings have not been described before for a site-specific analysis. Overall, this work provides a methodology for performing a dedicated site-specific search for sulfated and *O*-acetylated *N*-glycans that can be easily transferred, e.g., to human IgA derived from mucosal tissues, milk, or saliva. The future aim of this work is to use the explorative strategies described in this work to maximize the *N*-glycoproteomic analysis of other glycoproteins of relevance, to include rare *N*-glycans on its site-specific *N*-glycosylation profile.

Intentionally blank page

6 Conclusion and Outlook

Please note that parts of this section are taken from the original publications Zuniga-Banuelos et al. 2023, 2025a and Zuniga-Banuelos et al. 2024, 2025b [226-229].

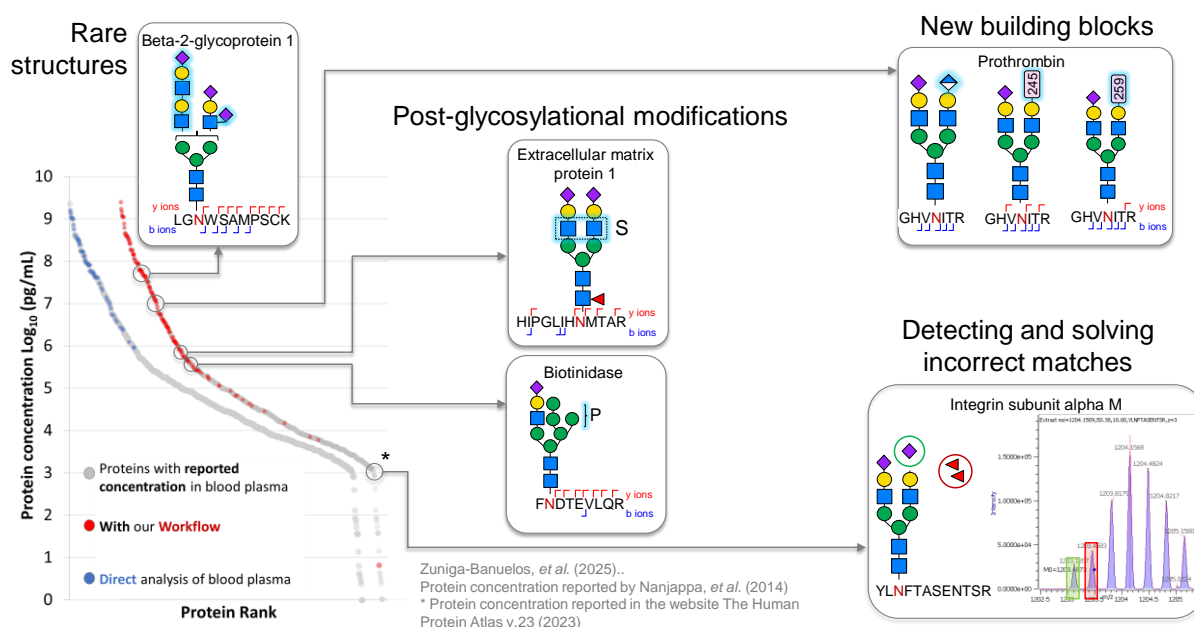


Figure 6.1. Gain obtained with the workflow developed for the in depth-N-glycoproteomic analysis of human blood plasma proteins.

The objectives of this thesis were i) to establish a workflow for the in-depth *N*-glycoproteomic analysis of human blood plasma proteins and ii) to apply the developed *N*-glycoproteomic workflow for the site-specific characterization of an *N*-glycan that has been challenging via current glycoproteomic methodologies.

This thesis addressed significant limitations of an in-depth glycoproteomic analysis in a complex sample, such as blood plasma. These limitations can be summarized in: 1) the challenging protein concentration range in blood plasma, 2) the high diversity of protein glycoforms, 3) the *N*-glycan structural complexity, and 4) the incomplete overview of the input spectra—common parameters can reliably deduce only a part of the gPSMs. Therefore, in the first project, a workflow for the in-depth *N*-glycoproteomic analysis of human blood plasma was developed and applied. This workflow comprises fractionation of HAP-depleted blood plasma and glycopeptide enrichment followed by two LC-MS/MS measurements using the fragmentation energies HCD.step and HCD.low. As displayed in Figure 6.1, the devised workflow enabled the detection of glycoproteins in a concentration range from 10^9 to 10^3 pg/mL, thus expanding the detection range by five orders of magnitude compared to the direct analysis of blood plasma. The potential of new *N*-glycoproteomic data search strategies was also explored. Validation and curation of the *N*-glycoproteomic search are based on a novel gPSM decision tree to critically assess evidence of the results of MS analyses on the peptide and *N*-glycan composition. This approach allowed us to reliably elucidate structural glycan features and identify rare *N*-glycan compositions. This included, for instance, the presence of glucuronic acid and two rare *N*-glycan building blocks (245.0524 and 259.0672 Da). Furthermore, an atypical disialylated-antenna *N*-glycan structure containing a HexNAc-NeuAc linkage was primarily detected based on the presence of the oxonium ion HexNAc₁NeuAc₁ [M+H]⁺. Although more analyses are required to confirm this finding, the elucidation of a NeuAc position in those *N*-glycopeptides would have been erroneously assumed to be spread on different antennae without HCD.low spectra annotation. As presented in this thesis, other structural groups that are hard to discriminate, such as antenna versus core fucose, diLacNAc versus multi-antennae, bisecting GlcNAc, and phosphorylation versus sulfation, were also reliably confirmed. In contrast to previous reports, no hexose sulfation was detected, but only HexNAc sulfation, with high certainty through oxonium marker ions. Using manual validation, this work contributes to uncovering problems, pitfalls, and opportunities in glycoproteomics research applied to biomarker discovery, biotherapeutic products, and biochemistry. In the future, the key insights learned from manual validation will support the establishment of a more efficient but reliable data analysis workflow for obtaining an accurate description of the *N*-glycoproteome of any sample.

The second project applied the multiple MS-based analyses reporting IgA *N*-glycoproteomics that overlooked sulfated and *O*-acetylated *N*-glycans despite their previous identification by glycomic analysis [28, 45, 186, 187]. IgA is an antibody with clinical relevance due to its multiple roles in the immune system.

In view of the importance of this immunoglobulin and its incomplete glycoproteomic analysis, the developed in-depth *N*-glycoproteomic workflow was applied. Two commercially available human IgA samples isolated from blood serum were implemented to compare the presence of sulfated and other rare *N*-glycans. Multiple search strategies were applied to maximize the identification of rare *N*-glycans using fractionated and unfractionated human serum IgA. This work achieved the identification of the *N*-glycosylation sites at the C_H2 domain and tailpiece harboring the sulfated *N*-glycan HexNAc₄Hex₅Fuc₁NeuAc₂Sulfo₁ (FA2G2S2-SO₄), reported before only by site unspecific *N*-glycomic experiments [44, 45]. Six other novel hybrid-type and complex-type HexNAc-sulfated *N*-glycans in IgA were site-specifically detected for the first time. These compositions had been overlooked in previous glycomic and glycoproteomic analyses [7, 110, 152, 183, 184, 290]. However, previously reported galactose-sulfated *N*-glycans in IgA were not detected [154, 186, 187]. In addition, new *N*-glycan compositions holding *O*-acetylated NeuAc were identified in a site-specific manner. Finally, the relative abundance of sulfated and *O*-acetylated *N*-glycans was estimated per glycosylation site and per IgA sample. The data generated demonstrate, for the first time, that the primary position in IgA for sulfated *N*-glycans is the tailpiece. It was detected that the relative abundance of sulfated *N*-glycans at the tailpiece *N*-glycosylation site is at least 10 times superior to the abundance of NeuAc *O*-acetylated *N*-glycans at the same site. Interestingly, NeuAc *O*-acetylated *N*-glycans were only identified in the IgA sample with the highest abundance of IgA2 subclass. Surprisingly *N*-glycopeptides from contaminant proteins bearing glucuronidated *N*-glycans were identified in both commercial IgA samples. The *N*-glycans linked to the tailpiece position of IgA play an important role in the protection against infection by sialic-acid-binding viruses [172], for the binding to complement C3, and in IgA dimerization [188]. Our MS-based *N*-glycoproteomic workflow allowed the site-specific investigation of very low abundant *N*-glycopeptide forms, such as HexNAc-sulfated and NeuAc *O*-acetylated *N*-glycans. Finally, this work also implemented an oxonium-ion-guided strategy for determination of parameters necessary to identify the *N*-glycopeptides of interest in IgA. This oxonium-ion-guided strategy can be adapted to other instances, and it should be included as part of any glycoproteomic data analysis workflows as a diagnostic step. Particularly the integration of oxonium-ion-guided *N*-glycan diagnostic in a software would support the identification of MS² spectra that require peptide or *N*-glycan components not initially present in the database.

Results obtained in this thesis encourages the application of the developed workflow to other biological samples, e.g., glycoproteins isolated from different body fluids, such as urine, milk, saliva, or mucosal secretions. For example, its application in longitudinal studies focused on evaluating the site-specific variation of low-abundant *N*-glycans in health and disease. In the future, using artificial intelligence in glycobiology could offer new opportunities, especially for speeding up data interpretation and validation [315]. On the one hand, the prediction of protein *N*-glycosylation and, on the other hand, the screening of

its (patho-) physiological role with the integration of multiple omics. For this purpose, retrieving and integrating all existing knowledge about protein *N*-glycosylation is crucial. Nevertheless, the bottleneck is that the current glycoproteomic bioinformatics resources still need to be increased. Glycoproteomic information linking glycosylation features to biological function must be included, and datasets from different analytical platforms and research works should be easier to obtain [315]. Improvements in bioinformatics resources can, therefore, significantly impact the field of glycobiology. In particular, it will allow a deeper understanding of the role of *N*-glycan modifications in complex biological systems, as well as a faster and more effective screening for glycoprotein-based therapies or biomarkers.

References

1. Reece, A.S., *Extending the “Paracentral Dogma” of Biology with the Metabolome: Implications for Understanding Genomic–Glycomic–Metabolic–Epigenomic Synchronization*. Engineering, 2023. **26**. DOI: 10.1016/j.eng.2022.07.011.
2. Marth, J.D., *A unified vision of the building blocks of life*. Nat. Cell. Biol., 2008. **10**(9): p. 1015-1016.DOI: 10.1038/ncb0908-1015.
3. Feichtinger, R.G., A. Hullen, A. Koller, D. Kotzot, et al., *A spoonful of L-fucose-an efficient therapy for GFUS-CDG, a new glycosylation disorder*. EMBO Mol. Med., 2021. **13**(9).DOI: 10.15252/emmm.202114332.
4. Sturiale, L., M.C. Nassogne, A. Palmigiano, A. Messina, et al., *Aberrant sialylation in a patient with a HNF1 alpha variant and liver adenomatosis*. Iscience, 2021. **24**(4). DOI: 10.1016/j.isci.2021.102323.
5. Kirwan, A., M. Utratna, M.E. O'Dwyer, L. Joshi, and M. Kilcoyne, *Glycosylation-Based Serum Biomarkers for Cancer Diagnostics and Prognostics*. Biomed. Res. Int., 2015. **2015**(490531).DOI: 10.1155/2015/490531.
6. Bondt, A., M.H.J. Selman, A.M. Deelder, J.M.W. Hazes, et al., *Association between Galactosylation of Immunoglobulin G and Improvement of Rheumatoid Arthritis during Pregnancy Is Independent of Sialylation*. J. Proteome. Res., 2013. **12**(10): p. 4522-4531.DOI: 10.1021/pr400589m.
7. Bondt, A., S. Nicolardi, B.C. Jansen, K. Stavenhagen, et al., *Longitudinal monitoring of immunoglobulin A glycosylation during pregnancy by simultaneous MALDI-FTICR-MS analysis of N- and O-glycopeptides*. Sci. Rep., 2016. **6**: p. 1-12.DOI: 10.1038/srep27955.
8. Deris, H., P. Tominac, F. Vuckovic, A. Astrup, et al., *Susceptibility of Human Plasma N-glycome to Low-Calorie and Different Weight-Maintenance Diets*. Int. J. Mol. Sci., 2022. **23**(24).DOI: 10.3390/ijms232415772.
9. Reiding, K.R., A. Bondt, R. Hennig, R.A. Gardner, et al., *High-throughput Serum N-Glycomics: Method Comparison and Application to Study Rheumatoid Arthritis and Pregnancy-associated Changes*. Mol. Cell. Proteomics., 2019. **18**(1): p. 3-15.DOI: 10.1074/mcp.RA117.000454.
10. Kristic, J., F. Vuckovic, C. Menni, L. Klaric, et al., *Glycans are a novel biomarker of chronological and biological ages*. J. Gerontol. A. Biol. Sci. Med. Sci., 2014. **69**(7): p. 779-789.DOI: 10.1093/gerona/glt190.
11. Parekh, R., I. Roitt, D. Isenberg, R. Dwek, and T. Rademacher, *Age-related galactosylation of the N-linked oligosaccharides of human serum IgG*. J. Exp. Med., 1988. **167**(5): p. 1731-1736.DOI: 10.1084/jem.167.5.1731.
12. Lombard, J., *The multiple evolutionary origins of the eukaryotic N-glycosylation pathway*. Biol. Direct., 2016. **11**(36).DOI: 10.1186/s13062-016-0137-2.
13. Moriya, S., M. Morimoto, K. Numata, A. Nozaki, et al., *Fucosylated Fraction of Alpha-fetoprotein as a Serological Marker of Early Hepatocellular Carcinoma*. Anticancer Res., 2013. **33**(3): p. 997-1001.
14. Cunningham, L.W., B.J. Nuenke, and R.B. Nuenke, *Preparation of glycopeptides from ovalbumin*. Biochim. Biophys. Acta, 1957. **26**(3): p. 660-661.DOI: 10.1016/0006-3002(57)90126-9.
15. Nuenke, R.H. and L.W. Cunningham, *Preparation and structural studies of ovalbumin glycopeptides*. J. Biol. Chem., 1961. **236**: p. 2452-2460.
16. Gludovacz, E., D. Maresch, L. Lopes de Carvalho, V. Puxbaum, et al., *Oligomannosidic glycans at Asn-110 are essential for secretion of human diamine oxidase*. J. Biol. Chem., 2018. **293**(3): p. 1070-1087.DOI: 10.1074/jbc.M117.814244.
17. Kellokumpu, S., *Golgi pH, Ion and Redox Homeostasis: How Much Do They Really Matter?* Front. Cell. Dev. Biol., 2019. **7**(93).DOI: 10.3389/fcell.2019.00093.
18. Hassinen, A., F. Khoder-Agha, E. Khosrowabadi, D. Mennerich, et al., *A Golgi-associated redox switch regulates catalytic activation and cooperative functioning of ST6Gal-I with B4GalT-I*. Redox Biol., 2019. **24**(101182).DOI: 10.1016/j.redox.2019.101182.
19. Zoldos, V., T. Horvat, M. Novokmet, C. Cuenin, et al., *Epigenetic silencing of HNF1A associates with changes in the composition of the human plasma N-glycome*. Epigenetics, 2012. **7**(2): p. 164-172.DOI: 10.4161/epi.7.2.18918.
20. Joshi, H.J., L. Hansen, Y. Narimatsu, H.H. Freeze, et al., *Glycosyltransferase genes that cause monogenic congenital disorders of glycosylation are distinct from glycosyltransferase genes associated with complex diseases*. Glycobiology, 2018. **28**(5): p. 284-294.DOI: 10.1093/glycob/cwy015.

21. Bernardi, C., U. Soffientini, F. Piacente, and M.G. Tonetti, *Effects of microRNAs on fucosyltransferase 8 (FUT8) expression in hepatocarcinoma cells*. PLoS One, 2013. **8**(10): p. e76540.DOI: 10.1371/journal.pone.0076540.
22. Pothukuchi, P., I. Agliarulo, D. Russo, R. Rizzo, et al., *Translation of genome to glycome: role of the Golgi apparatus*. FEBS Lett., 2019. **593**(17): p. 2390-2411.DOI: 10.1002/1873-3468.13541.
23. Au, C.E., L. Hermo, E. Byrne, J. Smirle, et al., *Expression, sorting, and segregation of Golgi proteins during germ cell differentiation in the testis*. Mol. Biol. Cell., 2015. **26**(22): p. 4015-4032.DOI: 10.1091/mbc.E14-12-1632.
24. Anderson, N.L. and N.G. Anderson, *The human plasma proteome: history, character, and diagnostic prospects*. Mol. Cell. Proteomics., 2002. **1**(11): p. 845-867.DOI: 10.1074/mcp.r200007-mcp200.
25. Nanjappa, V., J.K. Thomas, A. Marimuthu, B. Muthusamy, et al., *Plasma Proteome Database as a resource for proteomics research: 2014 update*. Nucleic Acids Res., 2014. **42**(D1): p. D959-D965.DOI: 10.1093/nar/gkt1251.
26. Wang, M., C.J. Herrmann, M. Simonovic, D. Szklarczyk, and C. von Mering, *Version 4.0 of PaxDb: Protein abundance data, integrated across model organisms, tissues, and cell-lines*. Proteomics, 2015. **15**(18): p. 3163-3168.DOI: 10.1002/pmic.201400441.
27. Cox, J., M.Y. Hein, C.A. Lubner, I. Paron, et al., *Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ*. Mol. Cell. Proteomics., 2014. **13**(9): p. 2513-2526.DOI: 10.1074/mcp.M113.031591.
28. Boisgard, R., G. Charpigny, and E. Chanut, *Polymeric IgA are sulfated proteins*. FEBS Lett., 1999. **463**(3): p. 250-254.DOI: 10.1016/S0014-5793(99)01631-2.
29. Yamada, K., K. Suzuki, Y. Hirohata, and M. Kinoshita, *Analysis of Minor Acidic N-Glycans in Human Serum*. J. Proteome Res., 2020. **19**(8): p. 3033-3043.DOI: 10.1021/acs.jproteome.0c00079.
30. Lee, P.Y., J. Osman, T.Y. Low, and R. Jamal, *Plasma/serum proteomics: depletion strategies for reducing high-abundance proteins for biomarker discovery*. Bioanalysis, 2019. **11**(19): p. 1799-1812.DOI: 10.4155/bio-2019-0145.
31. Kaur, G., A. Poljak, S.A. Ali, L. Zhong, et al., *Extending the Depth of Human Plasma Proteome Coverage Using Simple Fractionation Techniques*. J. Proteome Res., 2021. **20**(2): p. 1261-1279.DOI: 10.1021/acs.jproteome.0c00670.
32. Thiesler, C.T., S. Cajic, D. Hoffmann, C. Thiel, et al., *Glycomic Characterization of Induced Pluripotent Stem Cells Derived from a Patient Suffering from Phosphomannomutase 2 Congenital Disorder of Glycosylation (PMM2-CDG)*. Mol. Cell. Proteom., 2016. **15**(4): p. 1435-1452.DOI: 10.1074/mcp.M115.054122.
33. Beimdiek, J., R. Hennig, R. Burock, O. Puk, et al., *Serum N-glycomics of a novel CDG-IIb patient reveals aberrant IgG glycosylation*. Glycobiology, 2022. **32**(5): p. 380-390.DOI: 10.1093/glycob/cwac003.
34. Puan, K.J., B. San Luis, N. Yusof, D. Kumar, et al., *FUT6 deficiency compromises basophil function by selectively abrogating their sialyl-Lewis x expression*. Commun. Biol., 2021. **4**(1): p. 832.DOI: 10.1038/s42003-021-02295-8.
35. Parekh, R.B., R.A. Dwek, B.J. Sutton, D.L. Fernandes, et al., *Association of rheumatoid arthritis and primary osteoarthritis with changes in the glycosylation pattern of total serum IgG*. Nature, 1985. **316**(6027): p. 452-457.DOI: 10.1038/316452a0.
36. Flynn, G.C., X. Chen, Y.D. Liu, B. Shah, and Z. Zhang, *Naturally occurring glycan forms of human immunoglobulins G1 and G2*. Mol. Immunol., 2010. **47**(11-12): p. 2074-2082.DOI: 10.1016/j.molimm.2010.04.006.
37. Boune, S., P. Hu, A.L. Epstein, and L.A. Khawli, *Principles of N-Linked Glycosylation Variations of IgG-Based Therapeutics: Pharmacokinetic and Functional Considerations*. Antibodies, 2020. **9**(2): p. 22-41.DOI: 10.3390/antib9020022.
38. Martin, T.C., M. Simurina, M. Zabczynska, M. Martinic Kavur, et al., *Decreased Immunoglobulin G Core Fucosylation, A Player in Antibody-dependent Cell-mediated Cytotoxicity, is Associated with Autoimmune Thyroid Diseases*. Mol. Cell. Proteomics., 2020. **19**(5): p. 774-792.DOI: 10.1074/mcp.RA119.001860.
39. Cvetko, A., D. Kifer, O. Gornik, L. Klaric, et al., *Glycosylation Alterations in Multiple Sclerosis Show Increased Proinflammatory Potential*. Biomedicines, 2020. **8**(10): 410.DOI: 10.3390/biomedicines8100410.

40. Shinzaki, S., H. Iijima, T. Nakagawa, S. Egawa, et al., *IgG oligosaccharide alterations are a novel diagnostic marker for disease activity and the clinical course of inflammatory bowel disease*. *Am. J. Gastroenterol.*, 2008. **103**(5): p. 1173-1181.DOI: 10.1111/j.1572-0241.2007.01699.x.
41. Gudelj, I., P.P. Salo, I. Trbojevic-Akmacic, M. Albers, et al., *Low galactosylation of IgG associates with higher risk for future diagnosis of rheumatoid arthritis during 10 years of follow-up*. *Biochim. Biophys. Acta Mol. Basis. Dis.*, 2018. **1864**(6 Pt A): p. 2034-2039.DOI: 10.1016/j.bbadis.2018.03.018.
42. Egashira, Y., M. Suganuma, Y. Kataoka, Y. Higa, et al., *Establishment and characterization of a fucosylated alpha-fetoprotein-specific monoclonal antibody: a potential application for clinical research*. *Sci. Rep.*, 2019. **9**(12359). 12359.DOI: 10.1038/s41598-019-48821-x.
43. Suttapitugsakul, S., K. Stavenhagen, S. Donskaya, D.A. Bennett, et al., *Glycoproteomics Landscape of Asymptomatic and Symptomatic Human Alzheimer's Disease Brain*. *Mol. Cell. Proteom.*, 2022. **21**(12). 100433.DOI: 10.1016/j.mcpro.2022.100433.
44. Chuzel, L., S.L. Fossa, M.L. Boisvert, S. Cajic, et al., *Combining functional metagenomics and glycoanalytics to identify enzymes that facilitate structural characterization of sulfated N-glycans*. *Microb. Cell. Fact.*, 2021. **20**. 162.DOI: 10.1186/s12934-021-01652-w.
45. Cajic, S., R. Hennig, V. Grote, U. Reichl, and E. Rapp, *Removable Dyes—The Missing Link for In-Depth N-glycan Analysis via Multi-Method Approaches*. *Engineering*, 2023(23): p. 132-150.DOI: 10.1016/j.eng.2023.02.016.
46. Cummings, R.D., *The repertoire of glycan determinants in the human glycome*. *Mol. Biosyst.*, 2009. **5**(10): p. 1087-1104.DOI: 10.1039/b907931a.
47. Varki, A. and S. Kornfeld, *Historical Background and Overview*, in *Essentials of Glycobiology*, A. Varki, et al., Editors. 2022: Cold Spring Harbor (NY). p. 1-20.
48. Schnaar, R.L., R. Sandhoff, M. Tiemeyer, and T. Kinoshita, *Glycosphingolipids*, in *Essentials of Glycobiology*, A. Varki, et al., Editors. 2022: Cold Spring Harbor (NY). p. 129-140.
49. Komath, S.S., M. Fujita, G.W. Hart, M.A.J. Ferguson, and T. Kinoshita, *Glycosylphosphatidylinositol Anchors*, in *Essentials of Glycobiology*, A. Varki, et al., Editors. 2022: Cold Spring Harbor (NY). p. 141-154.
50. Merry, C.L.R., U. Lindahl, J. Couchman, and J.D. Esko, *Proteoglycans and Sulfated Glycosaminoglycans*, in *Essentials of Glycobiology*, A. Varki, et al., Editors. 2022: Cold Spring Harbor (NY). p. 217-232.
51. Doucey, M.A., D. Hess, R. Cacan, and J. Hofsteenge, *Protein C-mannosylation is enzyme-catalysed and uses dolichyl-phosphatide-mannose as a precursor*. *Mol. Biol. Cell.*, 1998. **9**(2): p. 291-300.DOI: DOI 10.1091/mbc.9.2.291.
52. Hofsteenge, J., K.G. Huwiler, B. Macek, D. Hess, et al., *C-mannosylation and O-fucosylation of the thrombospondin type 1 module*. *J. Biol. Chem.*, 2001. **276**(9): p. 6485-6498.DOI: DOI 10.1074/jbc.M008073200.
53. Stanley, P., K.W. Moremen, N.E. Lewis, N. Taniguchi, and M. Aebi, *N-Glycans*, in *Essentials of Glycobiology*, A. Varki, et al., Editors. 2022: Cold Spring Harbor (NY). p. 103-116.
54. Brockhausen, I. and P. Stanley, *O-GalNAc Glycans*, in *Essentials of Glycobiology*, A. Varki, et al., Editors. 2017: Cold Spring Harbor (NY).
55. Stanley, P. and R.D. Cummings, *Structures Common to Different Glycans*, in *Essentials of Glycobiology*, A. Varki, et al., Editors. 2015: Cold Spring Harbor (NY). p. 161-178.
56. Haltiwanger, R.S., L. Wells, H.H. Freeze, H. Jafar-Nejad, et al., *Other Classes of Eukaryotic Glycans*, in *Essentials of Glycobiology*, A. Varki, et al., Editors. 2022: Cold Spring Harbor (NY). p. 155-164.
57. Bloch, J.S., A. John, R. Mao, S. Mukherjee, et al., *Structure, sequon recognition and mechanism of tryptophan C-mannosyltransferase*. *Nat Chem Biol*, 2023. **19**(5): p. 575-584.DOI: 10.1038/s41589-022-01219-9.
58. Joshi, H.J., Y. Narimatsu, K.T. Schjoldager, H.L.P. Tytgat, et al., *SnapShot: O-Glycosylation Pathways across Kingdoms*. *Cell*, 2018. **172**(3): p. 632-632e2.DOI: 10.1016/j.cell.2018.01.016.
59. Chung, C.Y., N.I. Majewska, Q. Wang, J.T. Paul, and M.J. Betenbaugh, *SnapShot: N-Glycosylation Processing Pathways across Kingdoms*. *Cell*, 2017. **171**(1): p. 258-258 e1 DOI: 10.1016/j.cell.2017.09.014.
60. Valguarnera, E., R.L. Kinsella, and M.F. Feldman, *Sugar and Spice Make Bacteria Not Nice: Protein Glycosylation and Its Influence in Pathogenesis*. *J. Mol. Biol.*, 2016. **428**(16): p. 3206-3220.DOI: 10.1016/j.jmb.2016.04.013.
61. Freeze, H.H., G.W. Hart, and R.L. Schnaar, *Glycosylation Precursors*, in *Essentials of Glycobiology*, A. Varki, et al., Editors. 2022, Cold Spring Harbor (NY).

62. Seeberger, P.H., *Monosaccharide Diversity*, in *Essentials of Glycobiology*, A. Varki, et al., Editors. 2017: Cold Spring Harbor (NY).
63. Morise, J., H. Takematsu, and S. Oka, *The role of human natural killer-1 (HNK-1) carbohydrate in neuronal plasticity and disease*. *Biochim. Biophys. Acta Gen. Subj.*, 2017. **1861**(10): p. 2455-2461.DOI: 10.1016/j.bbagen.2017.06.025.
64. Lewis, A.L., X. Chen, R.L. Schnaar, and A. Varki, *Sialic Acids and Other Nonulosonic Acids*, in *Essentials of Glycobiology*, A. Varki, et al., Editors. 2022: Cold Spring Harbor (NY). p. 185-204.
65. Hermentin, P., *Carbohydrate analysis*, in *Bioanalytics: Analytical Methods and Concepts in Biochemistry and Molecular Biology*, F. Lottspeich and J.W. Engels, Editors. 2018, Wiley-VCH Verlag GmbH & Co. KGaA. p. 574-612.
66. van Kooyk, Y. and G.A. Rabinovich, *Protein-glycan interactions in the control of innate and adaptive immune responses*. *Nat. Immunol.*, 2008. **9**(6): p. 593-601.DOI: 10.1038/ni.f.203.
67. Dahms, N., T. Bräulke, and A. Varki, *P-Type Lectins*, in *Essentials of Glycobiology*, A. Varki, et al., Editors. 2022: Cold Spring Harbor (NY). p. 443-454.
68. Lindahl, U., J. Couchman, K. Kimata, and J.D. Esko, *Proteoglycans and Sulfated Glycosaminoglycans*, in *Essentials of Glycobiology*, A. Varki, et al., Editors. 2017: Cold Spring Harbor (NY). p. 443-454.
69. Stanley, P., M. Wuhrer, G. Lauc, S.R. Stowell, and R.D. Cummings, *Structures Common to Different Glycans*, in *Essentials of Glycobiology*, A. Varki, et al., Editors. 2022: Cold Spring Harbor (NY). p. 165-184.
70. Kawade, H., J. Morise, S.K. Mishra, S. Tsujioka, et al., *Tissue-Specific Regulation of HNK-1 Biosynthesis by Bisecting GlcNAc*. *Molecules*, 2021. **26**(17): 5176.DOI: 10.3390/molecules26175176.
71. Yamamoto, S., S. Oka, M. Inoue, M. Shimuta, et al., *Mice deficient in nervous system-specific carbohydrate epitope HNK-1 exhibit impaired synaptic plasticity and spatial learning*. *J. Biol. Chem.*, 2002. **277**(30): p. 27227-27231.DOI: 10.1074/jbc.C200296200.
72. Tagawa, H., Y. Kizuka, T. Ikeda, S. Itoh, et al., *A non-sulfated form of the HNK-1 carbohydrate is expressed in mouse kidney*. *J. Biol. Chem.*, 2005. **280**(25): p. 23876-23883.DOI: 10.1074/jbc.M501728200.
73. Gludovacz, E., D. Maresch, L.L. de Carvalho, V. Puxbaum, et al., *Oligomannosidic glycans at Asn-110 are essential for secretion of human diamine oxidase*. *J. Biol. Chem.*, 2018. **293**(3): p. 1070-1087.DOI: 10.1074/jbc.M117.814244.
74. Serrato, J.A., V. Hernandez, S. Estrada-Mondaca, L.A. Palomares, and O.T. Ramirez, *Differences in the glycosylation profile of a monoclonal antibody produced by hybridomas cultured in serum-supplemented, serum-free or chemically defined media*. *Biotechnol. Appl. Biochem.*, 2007. **47**(2): p. 113-124.DOI: 10.1042/BA20060216.
75. Pralow, A., M. Hoffmann, T. Nguyen-Khuong, M. Pioch, et al., *Comprehensive N-glycosylation analysis of the influenza A virus proteins HA and NA from adherent and suspension MDCK cells*. *FEBS J.*, 2021. **288**(16): p. 4869-4891.DOI: 10.1111/febs.15787.
76. Rombouts, Y., E. Ewing, L.A. van de Stadt, M.H.J. Selman, et al., *Anti-citrullinated protein antibodies acquire a pro-inflammatory Fc glycosylation phenotype prior to the onset of rheumatoid arthritis*. *Ann. Rheum. Dis.*, 2015. **74**(1): p. 234-241.DOI: 10.1136/annrheumdis-2013-203565.
77. Apweiler, R., H. Hermjakob, and N. Sharon, *On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database*. *Biochimica Et Biophysica Acta-General Subjects*, 1999. **1473**(1): p. 4-8.DOI: Doi 10.1016/S0304-4165(99)00165-8.
78. Petrescu, A.J., A.L. Milac, S.M. Petrescu, R.A. Dwek, and M.R. Wormald, *Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding*. *Glycobiology*, 2004. **14**(2): p. 103-114.DOI: 10.1093/glycob/cwh008.
79. Slynko, V., M. Schubert, S. Numao, M. Kowarik, et al., *NMR structure determination of a segmentally labeled glycoprotein using in vitro glycosylation*. *J. Am. Chem. Soc.*, 2009. **131**(3): p. 1274-1281.DOI: 10.1021/ja808682v.
80. Grabarics, M., M. Lettow, C. Kirschbaum, K. Greis, et al., *Mass Spectrometry-Based Techniques to Elucidate the Sugar Code*. *Chem. Rev.*, 2022. **122**(8): p. 7840-7908.DOI: 10.1021/acs.chemrev.1c00380.
81. Narimatsu, Y., H.J. Joshi, R. Nason, J. Van Coillie, et al., *An Atlas of Human Glycosylation Pathways Enables Display of the Human Glycome by Gene Engineered Cells*. *Mol. Cell.*, 2019. **75**(2): p. 394-407.DOI: 10.1016/j.molcel.2019.05.017.

82. Taylor, M.E., K. Drickamer, A. Imberty, Y. van Kooyk, et al., *Discovery and Classification of Glycan-Binding Proteins*, in *Essentials of Glycobiology*, A. Varki, et al., Editors. 2022: Cold Spring Harbor (NY). p. 375-386.
83. Raposo, C.D., A.B. Canelas, and M.T. Barros, *Human Lectins, Their Carbohydrate Affinities and Where to Find Them*. *Biomolecules*, 2021. **11**(2). 188.DOI: 10.3390/biom11020188.
84. Smith, B.A.H. and C.R. Bertozzi, *The clinical impact of glycobiology: targeting selectins, Siglecs and mammalian glycans*. *Nat. Rev. Drug. Discov.*, 2021. **20**(3): p. 217-243.DOI: 10.1038/s41573-020-00093-1.
85. Gabius, H.J., *How to Crack the Sugar Code*. *Folia. Biol.*, 2017. **63**(4): p. 121-131.
86. Ivetic, A., H.L. Hoskins Green, and S.J. Hart, *L-selectin: A Major Regulator of Leukocyte Adhesion, Migration and Signaling*. *Front. Immunol.*, 2019. **10**. 1068.DOI: 10.3389/fimmu.2019.01068.
87. Foster, G.A., R.M. Gower, K.L. Stanhope, P.J. Havel, et al., *On-chip phenotypic analysis of inflammatory monocytes in atherogenesis and myocardial infarction*. *Proc. Natl. Acad. Sci. U. S. A.*, 2013. **110**(34): p. 13944-13949.DOI: 10.1073/pnas.1300651110.
88. Thiel, M., C. Zourelidis, J.D. Chambers, U.H. vonAndrian, et al., *Expression of beta(2)-integrins and L-selectin on polymorphonuclear leukocytes in septic patients*. *Eur. Surg. Res.*, 1997. **29**(3): p. 160-175.DOI: Doi 10.1159/000129521.
89. Muthana, S.M., C.T. Campbell, and J.C. Gildersleeve, *Modifications of glycans: biological significance and therapeutic opportunities*. *ACS Chem. Biol.*, 2012. **7**(1): p. 31-43.DOI: 10.1021/cb2004466.
90. Rakus, J.F. and L.K. Mahal, *New technologies for glycomic analysis: toward a systematic understanding of the glycome*. *Annu. Rev. Anal. Chem.*, 2011. **4**: p. 367-392.DOI: 10.1146/annurev-anchem-061010-113951.
91. Lottspeich, F. and J.W. Engels, *Bioanalytics : analytical methods and concepts in biochemistry and molecular biology*. 2018: Wiley-VCH.
92. Larson, E.J., D.S. Roberts, J.A. Melby, K.M. Buck, et al., *High-Throughput Multi-attribute Analysis of Antibody-Drug Conjugates Enabled by Trapped Ion Mobility Spectrometry and Top-Down Mass Spectrometry*. *Anal. Chem.*, 2021. **93**(29): p. 10013-10021.DOI: 10.1021/acs.analchem.1c00150.
93. Zhang, Y., B.R. Fonslow, B. Shan, M.C. Baek, and J.R. Yates, 3rd, *Protein analysis by shotgun/bottom-up proteomics*. *Chem. Rev.*, 2013. **113**(4): p. 2343-2394.DOI: 10.1021/cr3003533.
94. Kolarich, D., P.H. Jensen, F. Altmann, and N.H. Packer, *Determination of site-specific glycan heterogeneity on glycoproteins*. *Nat. Protoc.*, 2012. **7**(7): p. 1285-1298.DOI: 10.1038/nprot.2012.062.
95. Rosenjack-Burchum, J., *Blood Cells and the Hematopoietic System*, in *Porth's Essentials of Pathophysiology*, T.L. Norris, Editor. 2020, Wolters Kluwer. p. 537-544.
96. Geyer, P.E., L.M. Holdt, D. Teupser, and M. Mann, *Revisiting biomarker discovery by plasma proteomics*. *Mol. Syst. Biol.*, 2017. **13**(9). 942.DOI: 10.15252/msb.20156297.
97. Million, R., S. Tolin, L. Puricelli, S. Sbrignadello, et al., *High abundance proteins depletion vs low abundance proteins enrichment: comparison of methods to reduce the plasma proteome complexity*. *PLoS One*, 2011. **6**(5).DOI: 10.1371/journal.pone.0019603.
98. Thaysen-Andersen, M., N.H. Packer, and B.L. Schulz, *Maturing Glycoproteomics Technologies Provide Unique Structural Insights into the N-glycoproteome and Its Regulation in Health and Disease*. *Mol. Cell. Proteom.*, 2016. **15**(6): p. 1773-1790.DOI: 10.1074/mcp.O115.057638.
99. Miura, Y., N. Hashii, Y. Ohta, Y. Itakura, et al., *Characteristic glycopeptides associated with extreme human longevity identified through plasma glycoproteomics*. *Biochim. Biophys. Acta - Gen. Subj.*, 2018. **1862**(6): p. 1462-1471.DOI: 10.1016/j.bbagen.2018.03.025.
100. Hennig, R., S. Cajic, M. Borowiak, M. Hoffmann, et al., *Towards personalized diagnostics via longitudinal study of the human plasma N-glycome*. *Biochim. Biophys. Acta - Gen. Subj.*, 2016. **1860**(8): p. 1728-1738.DOI: 10.1016/j.bbagen.2016.03.035.
101. Vreeker, G.C.M., M.R. Bladergroen, S. Nicolardi, W.E. Mesker, et al., *Dried blood spot N-glycome analysis by MALDI mass spectrometry*. *Talanta*, 2019. **205**(120104). 120104.DOI: 10.1016/j.talanta.2019.06.104.
102. Struwe, W.B., S. Agravat, K.F. Aoki-Kinoshita, M.P. Campbell, et al., *The minimum information required for a glycomics experiment (MIRAGE) project: sample preparation guidelines for reliable reporting of glycomics datasets*. *Glycobiology*, 2016. **26**(9): p. 907-910.DOI: 10.1093/glycob/cww082.
103. Campbell, M.P., J.L. Abrahams, E. Rapp, W.B. Struwe, et al., *The minimum information required for a glycomics experiment (MIRAGE) project: LC guidelines*. *Glycobiology*, 2019. **29**(5): p. 349-354.DOI: 10.1093/glycob/cwz009.

104. Kolarich, D., E. Rapp, W.B. Struwe, S.M. Haslam, et al., *The minimum information required for a glycomics experiment (MIRAGE) project: improving the standards for reporting mass-spectrometry-based glycoanalytic data*. Mol. Cell. Proteomics., 2013. **12**(4): p. 991-995.DOI: 10.1074/mcp.O112.026492.
105. Liu, Y., R. McBride, M. Stoll, A.S. Palma, et al., *The minimum information required for a glycomics experiment (MIRAGE) project: improving the standards for reporting glycan microarray-based data*. Glycobiology, 2017. **27**(4): p. 280-284.DOI: 10.1093/glycob/cww118.
106. Pongracz, T., M. Wuhler, and N. de Haan, *Expanding the Reaction Space of Linkage-Specific Sialic Acid Derivatization*. Molecules, 2019. **24**(19). 3617.DOI: 10.3390/molecules24193617.
107. Taylor, C.F., P.A. Binz, R. Aebersold, M. Affolter, et al., *Guidelines for reporting the use of mass spectrometry in proteomics*. Nat. Biotechnol., 2008. **26**(8): p. 860-861.DOI: 10.1038/nbt0808-860.
108. Deutsch, E.W., L. Lane, C.M. Overall, N. Bandeira, et al., *Human Proteome Project Mass Spectrometry Data Interpretation Guidelines 3.0*. J. Proteome Res., 2019. **18**(12): p. 4108-4116.DOI: 10.1021/acs.jproteome.9b00542.
109. Ramachandran, P., G. Xu, H.H. Huang, R. Rice, et al., *Serum Glycoprotein Markers in Nonalcoholic Steatohepatitis and Hepatocellular Carcinoma*. J. Proteome Res., 2022. **21**(4): p. 1083-1094.DOI: 10.1021/acs.jproteome.1c00965.
110. Plomp, R., N. de Haan, A. Bondt, J. Murli, et al., *Comparative Glycomics of Immunoglobulin A and G From Saliva and Plasma Reveals Biomarker Potential*. Front. Immunol., 2018. **9**(2436).DOI: 10.3389/fimmu.2018.02436.
111. Shu, Q., M. Li, L. Shu, Z. An, et al., *Large-scale Identification of N-linked Intact Glycopeptides in Human Serum using HILIC Enrichment and Spectral Library Search*. Mol. Cell. Proteomics., 2020. **19**(4): p. 672-689.DOI: 10.1074/mcp.RA119.001791.
112. Saraswat, M., K. Garapati, D.G. Mun, and A. Pandey, *Extensive heterogeneity of glycopeptides in plasma revealed by deep glycoproteomic analysis using size-exclusion chromatography*. Mol. Omics., 2021. **17**(6): p. 939-947.DOI: 10.1039/d1mo00132a.
113. Zhang, Y., Y. Mao, W. Zhao, T. Su, et al., *Glyco-CPLL: An Integrated Method for In-Depth and Comprehensive N-Glycoproteome Profiling of Human Plasma*. J. Proteome Res., 2020. **19**(2): p. 655-666.DOI: 10.1021/acs.jproteome.9b00557.
114. Keshishian, H., M.W. Burgess, M.A. Gillette, P. Mertins, et al., *Multiplexed, Quantitative Workflow for Sensitive Biomarker Discovery in Plasma Yields Novel Candidates for Early Myocardial Injury*. Mol. Cell. Proteomics., 2015. **14**(9): p. 2375-2393.DOI: 10.1074/mcp.M114.046813.
115. Varnavides, G., M. Madern, D. Anrather, N. Hartl, et al., *In Search of a Universal Method: A Comparative Survey of Bottom-Up Proteomics Sample Preparation Methods*. J. Proteome. Res., 2022. **21**(10): p. 2397-2411.DOI: 10.1021/acs.jproteome.2c00265.
116. Leon, I.R., V. Schwammle, O.N. Jensen, and R.R. Sprenger, *Quantitative assessment of in-solution digestion efficiency identifies optimal protocols for unbiased protein analysis*. Mol. Cell. Proteomics., 2013. **12**(10): p. 2992-3005.DOI: 10.1074/mcp.M112.025585.
117. Wisniewski, J.R., A. Zougman, N. Nagaraj, and M. Mann, *Universal sample preparation method for proteome analysis*. Nat. Methods, 2009. **6**(5): p. 359-362.DOI: 10.1038/nmeth.1322.
118. Hoffmann, M., M. Pioch, A. Pralow, R. Hennig, et al., *The Fine Art of Destruction: A Guide to In-Depth Glycoproteomic Analyses-Exploiting the Diagnostic Potential of Fragment Ions*. Proteomics, 2018. **18**(24). 1800282.DOI: 10.1002/pmic.201800282.
119. Bunkenborg, J., B.J. Pilch, A.V. Podtelejnikov, and J.R. Wisniewski, *Screening for N-glycosylated proteins by liquid chromatography mass spectrometry*. Proteomics, 2004. **4**(2): p. 454-465.DOI: 10.1002/pmic.200300556.
120. Oliveira, T., M. Thaysen-Andersen, N.H. Packer, and D. Kolarich, *The Hitchhiker's guide to glycoproteomics*. Biochem. Soc. Trans., 2021. **49**(4): p. 1643-1662.DOI: 10.1042/BST20200879.
121. Wan, H.H., J.Y. Yan, L. Yu, Q.Y. Sheng, et al., *Zirconia layer coated mesoporous silica microspheres as HILIC SPE materials for selective glycopeptide enrichment*. Analyst, 2011. **136**(21): p. 4422-4430.DOI: 10.1039/c1an15554g.
122. Snovidá, S.I., E.D. Bodnar, R. Viner, J. Saba, and H. Perreault, *A simple cellulose column procedure for selective enrichment of glycopeptides and characterization by nano LC coupled with electron-transfer and*

- high-energy collisional-dissociation tandem mass spectrometry*. Carbohydr. Res., 2010. **345**(6): p. 792-801.DOI: 10.1016/j.carres.2010.01.006.
123. Alagesan, K., S.K. Khilji, and D. Kolarich, *It is all about the solvent: on the importance of the mobile phase for ZIC-HILIC glycopeptide enrichment*. Analytical and Bioanalytical Chemistry, 2017. **409**(2): p. 529-538.DOI: 10.1007/s00216-016-0051-6.
 124. Selman, M.H., M. Hemayatkar, A.M. Deelder, and M. Wührer, *Cotton HILIC SPE microtips for microscale purification and enrichment of glycans and glycopeptides*. Anal. Chem., 2011. **83**(7): p. 2492-2499.DOI: 10.1021/ac1027116.
 125. Pradita, T., Y.J. Chen, E.G. Mernie, S.N. Bendulo, and Y.J. Chen, *ZIC-cHILIC Functionalized Magnetic Nanoparticle for Rapid and Sensitive Glycopeptide Enrichment from <1 microl Serum*. Nanomaterials, 2021. **11**(9).DOI: 10.3390/nano11092159.
 126. Buszewski, B. and S. Noga, *Hydrophilic interaction liquid chromatography (HILIC)-a powerful separation technique*. Analytical and Bioanalytical Chemistry, 2012. **402**(1): p. 231-247.DOI: 10.1007/s00216-011-5308-5.
 127. Zhang, Q., F.Q. Yang, L. Ge, Y.J. Hu, and Z.N. Xia, *Recent applications of hydrophilic interaction liquid chromatography in pharmaceutical analysis*. J. Sep. Sci., 2017. **40**(1): p. 49-80.DOI: 10.1002/jssc.201600843.
 128. Kozlik, P., R. Goldman, and M. Sanda, *Study of structure-dependent chromatographic behavior of glycopeptides using reversed phase nanoLC*. Electrophoresis, 2017. **38**(17): p. 2193-2199.DOI: 10.1002/elps.201600547.
 129. Chernykh, A., R. Kawahara, and M. Thaysen-Andersen, *Towards structure-focused glycoproteomics*. Biochem. Soc. Trans., 2021. **49**(1): p. 161-186.DOI: 10.1042/BST20200222.
 130. Wührer, M., C.A. Koeleman, C.H. Hokke, and A.M. Deelder, *Protein glycosylation analyzed by normal-phase nano-liquid chromatography--mass spectrometry of glycopeptides*. Anal. Chem., 2005. **77**(3): p. 886-894.DOI: 10.1021/ac048619x.
 131. Stavenhagen, K., R. Plomp, and M. Wührer, *Site-Specific Protein N- and O-Glycosylation Analysis by a C18-Porous Graphitized Carbon-Liquid Chromatography-Electrospray Ionization Mass Spectrometry Approach Using Pronase Treated Glycopeptides*. Anal. Chem., 2015. **87**(23): p. 11691-11699.DOI: 10.1021/acs.analchem.5b02366.
 132. Xiao, K. and Z. Tian, *GPSeeker Enables Quantitative Structural N-Glycoproteomics for Site- and Structure-Specific Characterization of Differentially Expressed N-Glycosylation in Hepatocellular Carcinoma*. J. Proteome Res., 2019. **18**(7): p. 2885-2895.DOI: 10.1021/acs.jproteome.9b00191.
 133. Sun, L., G. Zhu, and N.J. Dovichi, *Comparison of the LTQ-Orbitrap Velos and the Q-Exactive for proteomic analysis of 1-1000 ng RAW 264.7 cell lysate digests*. Rapid Commun. Mass Spectrom., 2013. **27**(1): p. 157-162.DOI: 10.1002/rcm.6437.
 134. Thermo Fisher Scientific Inc. *Orbitrap Eclipse™ Tribrid™ Mass Spectrometer*. 2023 [cited 2023 19.07.2023]; Available from: <https://www.thermofisher.com/order/catalog/product/de/en/FSN04-10000>.
 135. Vekey, K., O. Ozohanics, E. Toth, A. Jeko, et al., *Fragmentation characteristics of glycopeptides*. International Journal of Mass Spectrometry, 2013. **345**: p. 71-79.DOI: 10.1016/j.ijms.2012.08.031.
 136. Riley, N.M., S.A. Malaker, M.D. Driessen, and C.R. Bertozzi, *Optimal Dissociation Methods Differ for N- and O-Glycopeptides*. J. Proteome Res., 2020. **19**(8): p. 3286-3301.DOI: 10.1021/acs.jproteome.0c00218.
 137. Zeng, W., S. Zheng, T. Su, J. Cheng, et al., *Comparative N-Glycoproteomics Analysis of Clinical Samples Via Different Mass Spectrometry Dissociation Methods*. Front. Chem., 2022. **10**(839470).DOI: 10.3389/fchem.2022.839470.
 138. Kuo, C.W., S.Y. Guu, and K.H. Khoo, *Distinctive and Complementary MS2 Fragmentation Characteristics for Identification of Sulfated Sialylated N-Glycopeptides by nanoLC-MS/MS Workflow*. Journal of the American Society for Mass Spectrometry, 2018. **29**(6): p. 1166-1178.DOI: 10.1007/s13361-018-1919-9.
 139. Sanda, M., J. Benicky, and R. Goldman, *Low Collision Energy Fragmentation in Structure-Specific Glycoproteomics Analysis*. Anal. Chem., 2020. **92**(12): p. 8262-8267.DOI: 10.1021/acs.analchem.0c00519.
 140. Domon, B. and C.E. Costello, *A Systematic Nomenclature for Carbohydrate Fragmentations in Fab-MS Mass Spectra of Glycoconjugates*. Glycoconj. J., 1988. **5**(4): p. 397-409.DOI: 10.1007/Bf01049915.
 141. Stadlmann, J., J. Taubenschmid, D. Wenzel, A. Gattlinger, et al., *Comparative glycoproteomics of stem cells identifies new players in ricin toxicity*. Nature, 2017. **549**(7673): p. 538-542.DOI: 10.1038/nature24015.

142. Wuhrer, M., M.I. Catalina, A.M. Deelder, and C.H. Hokke, *Glycoproteomics based on tandem mass spectrometry of glycopeptides*. J. Chromatogr. B Analyt. Technol. Biomed. Life Sci., 2007. **849**(1-2): p. 115-128.DOI: 10.1016/j.jchromb.2006.09.041.
143. Dong, Q., X. Yan, Y. Liang, and S.E. Stein, *In-Depth Characterization and Spectral Library Building of Glycopeptides in the Tryptic Digest of a Monoclonal Antibody Using 1D and 2D LC-MS/MS*. J. Proteome Res., 2016. **15**(5): p. 1472-1486.DOI: 10.1021/acs.jproteome.5b01046.
144. Manz, C., M. Mancera-Arteu, A. Zappe, E. Hanozin, et al., *Determination of Sialic Acid Isomers from Released N-Glycans Using Ion Mobility Spectrometry*. Anal. Chem., 2022. **94**(39): p. 13323-13331.DOI: 10.1021/acs.analchem.2c00783.
145. Shen, J., L. Jia, L. Dang, Y. Su, et al., *StrucGP: de novo structural sequencing of site-specific N-glycan on glycoproteins using a modularization strategy*. Nat. Methods, 2021. **18**(8): p. 921-929.DOI: 10.1038/s41592-021-01209-0.
146. Cao, W.Q., M.Q. Liu, S.Y. Kong, M.X. Wu, et al., *Recent Advances in Software Tools for More Generic and Precise Intact Glycopeptide Analysis*. Mol. Cell. Proteomics., 2021. **20**. 100060.DOI: 10.1074/mcp.R120.002090.
147. Liu, T., W.J. Qian, M.A. Gritsenko, D.G. Camp, 2nd, et al., *Human plasma N-glycoproteome analysis by immunoaffinity subtraction, hydrazide chemistry, and mass spectrometry*. J. Proteome Res., 2005. **4**(6): p. 2070-2080.DOI: 10.1021/pr0502065.
148. Kawahara, R., A. Chernykh, K. Alagesan, M. Bern, et al., *Community evaluation of glycoproteomics informatics solutions reveals high-performance search strategies for serum glycopeptide analysis*. Nat. Methods, 2021. **18**(11): p. 1304-1316.DOI: 10.1038/s41592-021-01309-x.
149. Pioch, M., M. Hoffmann, A. Pralow, U. Reichl, and E. Rapp, *glyXtool(MS): An Open-Source Pipeline for Semiautomated Analysis of Glycopeptide Mass Spectrometry Data*. Anal. Chem., 2018. **90**(20): p. 11908-11916.DOI: 10.1021/acs.analchem.8b02087.
150. Hackett, W.E. and J. Zaia, *The Need for Community Standards to Enable Accurate Comparison of Glycoproteomics Algorithm Performance*. Molecules, 2021. **26**(16).DOI: 10.3390/molecules26164757.
151. Lee, L.Y., E.S. Moh, B.L. Parker, M. Bern, et al., *Toward Automated N-Glycopeptide Identification in Glycoproteomics*. J. Proteome Res., 2016. **15**(10): p. 3904-3915.DOI: 10.1021/acs.jproteome.6b00438.
152. Chandler, K.B., N. Mehta, D.R. Leon, T.J. Suscovich, et al., *Multi-isotype Glycoproteomic Characterization of Serum Antibody Heavy Chains Reveals Isotype- and Subclass-Specific N-Glycosylation Profiles*. Mol. Cell. Proteomics., 2019. **18**(4): p. 686-703.DOI: 10.1074/mcp.RA118.001185.
153. Huffman, J.E., A. Knezevic, V. Vitart, J. Kattla, et al., *Polymorphisms in B3GAT1, SLC9A9 and MGAT5 are associated with variation within the human plasma N-glycome of 3533 European adults*. Hum. Mol. Genet., 2011. **20**(24): p. 5000-5011.DOI: 10.1093/hmg/ddr414.
154. Alagesan, K., D.V. Silva, P.H. Seeberger, and D. Kolarich, *A novel, ultrasensitive approach for quantitative carbohydrate composition and linkage analysis using LC-ESI ion trap tandem mass spectrometry*. 2019. bioRxiv, DOI: 10.1101/853036.
155. Lei, M., M.V. Novotny, and Y. Mechref, *Sequential enrichment of sulfated glycans by strong anion-exchange chromatography prior to mass spectrometric measurements*. J. Am. Soc. Mass Spectrom., 2010. **21**(3): p. 348-357.DOI: 10.1016/j.jasms.2009.09.017.
156. Roushan, A., G.M. Wilson, D. Kletter, K.I. Sen, et al., *Peak Filtering, Peak Annotation, and Wildcard Search for Glycoproteomics*. Mol. Cell. Proteomics., 2021. **20**(100011).DOI: 10.1074/mcp.RA120.002260.
157. Zabczynska, M., P. Link-Lenczowski, and E. Pochech, *Glycosylation in Autoimmune Diseases*. Adv. Exp. Med. Biol., 2021. **1325**: p. 205-218.DOI: 10.1007/978-3-030-70115-4_10.
158. Vuckovic, F., J. Kristic, I. Gudelj, M. Teruel, et al., *Association of Systemic Lupus Erythematosus With Decreased Immunosuppressive Potential of the IgG Glycome*. Arthritis Rheumatol., 2015. **67**(11): p. 2978-2989.DOI: 10.1002/art.39273.
159. Trbojevic Akmacic, I., N.T. Ventham, E. Theodoratou, F. Vuckovic, et al., *Inflammatory bowel disease associates with proinflammatory potential of the immunoglobulin G glycome*. Inflamm. Bowel. Dis., 2015. **21**(6): p. 1237-1247.DOI: 10.1097/MIB.0000000000000372.
160. Alavi, A. and J.S. Axford, *Glyco-biomarkers: potential determinants of cellular physiology and pathology*. Dis. Markers, 2008. **25**(4-5): p. 193-205.DOI: 10.1155/2008/863032.

161. Cheng, Y., M. Li, S. Wang, H. Peng, et al., *Carbohydrate biomarkers for future disease detection and treatment*. Sci. China Chem., 2010. **53**(1): p. 3-20.DOI: 10.1007/s11426-010-0021-3.
162. Zhou, Q. and H. Qiu, *The Mechanistic Impact of N-Glycosylation on Stability, Pharmacokinetics, and Immunogenicity of Therapeutic Proteins*. J. Pharm. Sci., 2019. **108**(4): p. 1366-1377.DOI: 10.1016/j.xphs.2018.11.029.
163. Ruhaak, L.R., H.W. Uh, M. Beekman, C.H. Hokke, et al., *Plasma protein N-glycan profiles are associated with calendar age, familial longevity and health*. J. Proteome. Res., 2011. **10**(4): p. 1667-1674.DOI: 10.1021/pr1009959.
164. Knezevic, A., O. Gornik, O. Polasek, M. Pucic, et al., *Effects of aging, body mass index, plasma lipid profiles, and smoking on human plasma N-glycans*. Glycobiology, 2010. **20**(8): p. 959-969.DOI: 10.1093/glycob/cwq051.
165. Bondt, A., S. Nicolardi, B.C. Jansen, T.M. Kuijper, et al., *IgA N- and O-glycosylation profiling reveals no association with the pregnancy-related improvement in rheumatoid arthritis*. Arthritis. Res. Ther., 2017. **19**. 160.DOI: 10.1186/s13075-017-1367-0.
166. Shields, R.L., J. Lai, R. Keck, L.Y. O'Connell, et al., *Lack of fucose on human IgG1 N-linked oligosaccharide improves binding to human Fc gamma RIII and antibody-dependent cellular toxicity*. J. Biol. Chem., 2002. **277**(30): p. 26733-26740.DOI: 10.1074/jbc.M202069200.
167. Wada, R., M. Matsui, and N. Kawasaki, *Influence of N-glycosylation on effector functions and thermal stability of glycoengineered IgG1 monoclonal antibody with homogeneous glycoforms*. MAbs, 2019. **11**(2): p. 350-372.DOI: 10.1080/19420862.2018.1551044.
168. Ricklin, D., G. Hajishengallis, K. Yang, and J.D. Lambris, *Complement: a key system for immune surveillance and homeostasis*. Nat. Immunol., 2010. **11**(9): p. 785-797.DOI: 10.1038/ni.1923.
169. Son, M., B. Diamond, and F. Santiago-Schwarz, *Fundamental role of C1q in autoimmunity and inflammation*. Immunol. Res., 2015. **63**: p. 101-106.DOI: 10.1007/s12026-015-8705-6.
170. Ackerman, M.E., M. Crispin, X. Yu, K. Baruah, et al., *Natural variation in Fc glycosylation of HIV-specific antibodies impacts antiviral activity*. J. Clin. Invest., 2013. **123**(5): p. 2183-2192.DOI: 10.1172/JCI65708.
171. Kaneko, Y., F. Nimmerjahn, and J.V. Ravetch, *Anti-inflammatory activity of immunoglobulin G resulting from Fc sialylation*. Science, 2006. **313**(5787): p. 670-673.DOI: 10.1126/science.1129594.
172. Maurer, M.A., L. Meyer, M. Bianchi, H.L. Turner, et al., *Glycosylation of Human IgA Directly Inhibits Influenza A and Other Sialic-Acid-Binding Viruses*. Cell. Rep., 2018. **23**(1): p. 90-99.DOI: 10.1016/j.celrep.2018.03.027.
173. Huls, G., I.A. Heijnen, E. Cuomo, J. van der Linden, et al., *Antitumor immune effector mechanisms recruited by phage display-derived fully human IgG1 and IgA1 monoclonal antibodies*. Cancer Res., 1999. **59**(22): p. 5778-5784.
174. Monteiro, R.C., *The Role of IgA and IgA Fc Receptors as Anti-Inflammatory Agents*. J. Clin. Immunol., 2010. **30**: p. 61-64.DOI: 10.1007/s10875-010-9397-2.
175. Jorgensen, G.H., A. Gardulf, M.I. Sigurdsson, S.T. Sigurdardottir, et al., *Clinical symptoms in adults with selective IgA deficiency: a case-control study*. J. Clin. Immunol., 2013. **33**(4): p. 742-747.DOI: 10.1007/s10875-012-9858-x.
176. Takeuchi, T. and H. Ohno, *IgA in human health and diseases: Potential regulator of commensal microbiota*. Front. Immunol., 2022. **13**(1024330).DOI: 10.3389/fimmu.2022.1024330.
177. Ruiz-Palacios, G.M., J.J. Calva, L.K. Pickering, Y. Lopez-Vidal, et al., *Protection of breast-fed infants against Campylobacter diarrhea by antibodies in human milk*. J. Pediatr., 1990. **116**(5): p. 707-713.DOI: 10.1016/s0022-3476(05)82652-6.
178. Kerr, M.A., *The structure and function of human IgA*. Biochem. J., 1990. **271**(2): p. 285-296.DOI: 10.1042/bj2710285.
179. Mestecky, J., C. Lue, A. Tarkowski, I. Ladjeva, et al., *Comparative-Studies of the Biological Properties of Human-IgA Subclasses*. Protides Biol. Fluids., 1989. **36**: p. 173-182.
180. Torano, A. and F.W. Putnam, *Complete amino acid sequence of the alpha 2 heavy chain of a human IgA2 immunoglobulin of the A2m (2) allotype*. Proc. Natl. Acad. Sci. USA, 1978. **75**(2): p. 966-969.DOI: 10.1073/pnas.75.2.966.
181. Tsuzukida, Y., C.C. Wang, and F.W. Putnam, *Structure of the A2m(1) allotype of human IgA--a recombinant molecule*. Proc. Natl. Acad. Sci. USA, 1979. **76**(3): p. 1104-1108.DOI: 10.1073/pnas.76.3.1104.

182. Chintalacharuvu, K.R., M. Raines, and S.L. Morrison, *Divergence of human alpha-chain constant region gene sequences. A novel recombinant alpha 2 gene*. J. Immunol., 1994. **152**(11): p. 5299-5304.
183. Clerc, F., K.R. Reiding, N. de Haan, C.A.M. Koeleman, et al., *Immunoglobulin A Glycosylation Differs between Crohn's Disease and Ulcerative Colitis*. J. Proteome Res., 2023. **Article ASAP**.DOI: 10.1021/acs.jproteome.3c00260.
184. Steffen, U., C.A. Koeleman, M.V. Sokolova, H. Bang, et al., *IgA subclasses have different effector functions associated with distinct glycosylation profiles*. Nat. Commun., 2020. **11**(120). DOI: 10.1038/s41467-019-13992-8.
185. Shi, X., Y. Huang, Y. Mao, H. Naimy, and J. Zaia, *Tandem mass spectrometry of heparan sulfate negative ions: sulfate loss patterns and chemical modification methods for improvement of product ion profiles*. J. Am. Soc. Mass Spectrom., 2012. **23**(9): p. 1498-1511.DOI: 10.1007/s13361-012-0429-4.
186. GlyConnect. *N-glycan structures from Immunoglobulin heavy constant alpha 1*. 2019 [cited 2023 Jan-20-2023]; Available from: <https://glyconnect.expasy.org/browser/proteins/230>.
187. GlyConnect. *N-glycan structures from Immunoglobulin heavy constant alpha 2*. 2019 [cited 2023 Jan-20-2023]; Available from: <https://glyconnect.expasy.org/browser/proteins/647>.
188. Chuang, P.D. and S.L. Morrison, *Elimination of N-linked glycosylation sites from the human IgA1 constant region: effects on structure and function*. J. Immunol., 1997. **158**(2): p. 724-732.
189. Rifai, A., K. Fadden, S.L. Morrison, and K.R. Chintalacharuvu, *The N-glycans determine the differential blood clearance and hepatic uptake of human immunoglobulin (Ig)A1 and IgA2 isotypes*. J. Exp. Med., 2000. **191**(12): p. 2171-2182.DOI: 10.1084/jem.191.12.2171.
190. Reily, C., T.J. Stewart, M.B. Renfrow, and J. Novak, *Glycosylation in health and disease*. Nat. Rev. Nephrol., 2019. **15**(6): p. 346-366.DOI: 10.1038/s41581-019-0129-4.
191. Diseases, G.B.D. and C. Injuries, *Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019*. Lancet, 2020. **396**(10258): p. 1204-1222.DOI: 10.1016/S0140-6736(20)30925-9.
192. Global Burden of Disease (GBD) study. *GBD Compare: Alzheimer's disease and other dementias as cause of death or injury*. 2019 [cited 2023 07/29]; All ages, both sexes, highlighting Germany, comparing incidence rate in 1999 versus 2019.]. Available from: <https://vizhub.healthdata.org/gbd-compare/#0>.
193. Global Burden of Disease (GBD) study. *GBD Compare: Rheumatoid arthritis as cause of death or injury*. 2019 [cited 2023 07/29]; All ages, both sexes, highlighting Germany, comparing incidence rate in 1999 versus 2019.]. Available from: <https://vizhub.healthdata.org/gbd-compare/#0>.
194. World Health Organization (WHO). *The top 10 causes of death*. 2020 [cited 2023 07/29]; Available from: <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>.
195. Global Burden of Disease (GBD) study. *GBD Compare: Pancreatic cancer as a cause of death or injury*. 2019 [cited 2023 07/29]; B1.9. Pancreatic cancer. All ages, both sexes, highlighting Germany, comparing incidence rate in 1999 versus 2019.]. Available from: <https://vizhub.healthdata.org/gbd-compare/#0>.
196. Bratulic, S., F. Gatto, and J. Nielsen, *The Translational Status of Cancer Liquid Biopsies*. Regenerative Engineering and Translational Medicine, 2021. **7**(3): p. 312-352.DOI: 10.1007/s40883-019-00141-2.
197. Rasmussen, J. and H. Langerman, *Alzheimer's Disease - Why We Need Early Diagnosis*. Degenerative Neurological and Neuromuscular Disease, 2019. **9**: p. 123-130.DOI: 10.2147/Dnnd.S228939.
198. Hausmann, J.S., K.G. Lomax, A. Shapiro, and K. Durrant, *The patient journey to diagnosis and treatment of autoinflammatory diseases*. Orphanet. J. Rare Dis., 2018. **13**(156).DOI: 10.1186/s13023-018-0902-7.
199. Gopalakrishnan, S. and H. Ellinas, *Disorders of the Immune Response, Including HIV/AIDS*, in *Porth's Essentials of Pathophysiology*, T.L. Norris, Editor. 2020, Wolters Kluwer. p. 286-289.
200. Le Goueff, A., G. Smits, M. Delaunoy, I. Vandernoot, and F. Vandergheynst, *Genetic testing in autoinflammatory diseases - past, current and future perspectives*. Eur. J. Intern. Med., 2022. **106**: p. 71-79.DOI: 10.1016/j.ejim.2022.08.020.
201. Castro, C. and M. Gourley, *Diagnostic testing and interpretation of tests for autoimmunity*. J. Allergy. Clin. Immunol., 2010. **125**(2 Suppl 2): p. 238-247.DOI: 10.1016/j.jaci.2009.09.041.
202. Kissel, T., R.E.M. Toes, T.W.J. Huizinga, and M. Wuhler, *Glycobiology of rheumatic diseases*. Nat. Rev. Rheumatol., 2023. **19**(1): p. 28-43.DOI: 10.1038/s41584-022-00867-4.
203. Engdahl, C., A. Bondt, U. Harre, J. Raufer, et al., *Estrogen induces St6gal1 expression and increases IgG sialylation in mice and patients with rheumatoid arthritis: a potential explanation for the increased risk of*

- rheumatoid arthritis in postmenopausal women*. *Arthritis Res. Ther.*, 2018. **20**(84).DOI: 10.1186/s13075-018-1586-z.
204. Schwab, I. and F. Nimmerjahn, *Intravenous immunoglobulin therapy: how does IgG modulate the immune system?* *Nat. Rev. Immunol.*, 2013. **13**(3): p. 176-189.DOI: 10.1038/nri3401.
 205. Alves, I., M.M. Vicente, A.M. Dias, J. Gaifem, et al., *The Role of Glycosylation in Inflammatory Diseases*. *Adv Exp Med Biol*, 2021. **1325**: p. 265-283.DOI: 10.1007/978-3-030-70115-4_13.
 206. Araujo, L., P. Khim, H. Mkhikian, C.L. Mortales, and M. Demetriou, *Glycolysis and glutaminolysis cooperatively control T cell function by limiting metabolite supply to N-glycosylation*. *Elife*, 2017. **6**.DOI: 10.7554/eLife.21330.
 207. Wang, Y.L., A. Khan, A. Antonopoulos, L. Bouche, et al., *Loss of alpha 2-6 sialylation promotes the transformation of synovial fibroblasts into a pro-inflammatory phenotype in arthritis*. *Nat. Commun.*, 2021. **12**(2343). 2343.DOI: 10.1038/s41467-021-22365-z.
 208. Piedade, A., R. Francisco, J. Jaeken, P. Sarkhail, et al., *Epidemiology of congenital disorders of glycosylation (CDG)—overview and perspectives*. *J. Rare Dis.*, 2022. **1**(3).DOI: 10.1007/s44162-022-00003-6.
 209. Global Burden of Disease (GBD) study. *GBD Compare: Congenital birth defects as cause of death or injury*. 2019 [cited 2023 07/29]; All ages, both sexes, highlighting Germany, comparing incidence rate in 1999 versus 2019.]. Available from: <https://vizhub.healthdata.org/gbd-compare/#0>.
 210. Peanne, R., P. de Lonlay, F. Foulquier, U. Kornak, et al., *Congenital disorders of glycosylation (CDG): Quo vadis?* *Eur. J. Med. Genet.*, 2018. **61**(11): p. 643-663.DOI: 10.1016/j.ejmg.2017.10.012.
 211. Morticelli, L., C. Rossdam, S. Cajic, D. Böthig, et al., *Genetic knockout of porcine GGTA1 or CMAH/GGTA1 is associated with the emergence of neo-glycans*. *Transplantation*, 2023. **107**(10). e12804.DOI: 10.1097/01.tp.0000994748.86988.d3.
 212. Ng, B.G. and H.H. Freeze, *Perspectives on Glycosylation and Its Congenital Disorders*. *Trends Genet.*, 2018. **34**(6): p. 466-476.DOI: 10.1016/j.tig.2018.03.002.
 213. Losfeld, M.E., B.G. Ng, M. Kircher, K.J. Buckingham, et al., *A new Congenital Disorder of Glycosylation caused by a mutation in SSR4, the signal sequence receptor 4 protein of the TRAP complex*. *Glycobiology*, 2013. **23**(11): p. 1370-1370.
 214. Rickard, M. and A. Tomita-Mitchell, *Neoplasia*, in *Porth's Essentials of Pathophysiology*, T.L. Norris, Editor. 2020, Wolters Kluwer. p. 95-105.
 215. Hazelton, W.D. and E.G. Luebeck, *Biomarker-Based Early Cancer Detection: Is It Achievable?* *Sci. Transl. Med.*, 2011. **3**(109). 109fs9.DOI: 10.1126/scitranslmed.3003272.
 216. Miyoshi, E., K. Noda, Y. Yamaguchi, S. Inoue, et al., *The alpha1-6-fucosyltransferase gene and its biological significance*. *Biochim. Biophys. Acta*, 1999. **1473**(1): p. 9-20.DOI: 10.1016/s0304-4165(99)00166-x.
 217. Bastian, K., E. Scott, D.J. Elliott, and J. Munkley, *FUT8 Alpha-(1,6)-Fucosyltransferase in Cancer*. *Int J Mol Sci*, 2021. **22**(1).DOI: 10.3390/ijms22010455.
 218. Lise, M., C. Belluco, S.P. Perera, R. Patel, et al., *Clinical correlations of alpha2,6-sialyltransferase expression in colorectal cancer patients*. *Hybridoma*, 2000. **19**(4): p. 281-286.DOI: 10.1089/027245700429828.
 219. Holmes, E.H., G.K. Ostrander, H. Clausen, and N. Graem, *Oncofetal expression of Lex carbohydrate antigens in human colonic adenocarcinomas. Regulation through type 2 core chain synthesis rather than fucosylation*. *J. Biol. Chem.*, 1987. **262**(23): p. 11331-11338.
 220. Pinho, S.S. and C.A. Reis, *Glycosylation in cancer: mechanisms and clinical implications*. *Nat. Rev. Cancer*, 2015. **15**(9): p. 540-555.DOI: 10.1038/nrc3982.
 221. Alvarez, M.R.S., Q. Zhou, J. Tena, C.B. Lebrilla, et al., *N-Glycan and Glycopeptide Serum Biomarkers in Philippine Lung Cancer Patients Identified Using Liquid Chromatography-Tandem Mass Spectrometry*. *ACS Omega*, 2022. **7**(44): p. 40230-40240.DOI: 10.1021/acsomega.2c05111.
 222. Kontro, H., S. Joenvaara, C. Haglund, and R. Renkonen, *Comparison of sialylated N-glycopeptide levels in serum of pancreatic cancer patients, acute pancreatitis patients, and healthy controls*. *Proteomics*, 2014. **14**(15): p. 1713-1723.DOI: 10.1002/pmic.201300270.
 223. Hanna-Sawires, R.G., J.H. Schiphuis, M. Wuhler, H.F.A. Vasen, et al., *Clinical Perspective on Proteomic and Glycomic Biomarkers for Diagnosis, Prognosis, and Prediction of Pancreatic Cancer*. *Int. J. Mol. Sci.*, 2021. **22**(5). 2655.DOI: 10.3390/ijms22052655.
 224. Long, E.M. and F.W. Cao, *Disorders of Thought, Emotion, and Memory*, in *Porth's Essentials of Pathophysiology*, T.L. Norris, Editor. 2020, Wolters Kluwer. p. 470-472.

225. Garcia-Ayllon, M.S., A. Botella-Lopez, I. Cuchillo-Ibanez, A. Rabano, et al., *HNK-1 Carrier Glycoproteins Are Decreased in the Alzheimer's Disease Brain*. Mol. Neurobiol., 2017. **54**(1): p. 188-199.DOI: 10.1007/s12035-015-9644-x.
226. Zuniga-Banuelos, F.J., M. Hoffmann, U. Reichl, and E. Rapp, *New avenues for human blood plasma biomarker discovery via improved in-depth analysis of the low-abundant N-glycoproteome*. 2023. bioRxiv, 1-38 DOI: 10.1101/2023.10.22.562384.
227. Zuniga-Banuelos, F.J., G. Lemke, M. Hoffmann, U. Reichl, and E. Rapp, *Immunoglobulin A Carries Sulfated N-glycans Primarily at the Tailpiece Site –An Oxonium-Ion-Guided Approach for Site-Specific N-glycan Identification*. 2024. bioRxiv, DOI: 10.1101/2024.06.06.597690.
228. Zuniga-Banuelos, F.J., M. Hoffmann, U. Reichl, and E. Rapp, *New Avenues for Human Blood Plasma Biomarker Discovery via Improved In-Depth Analysis of the Low-Abundant N-glycoproteome*. Engineering, 2025.DOI: 10.1016/j.eng.2024.11.039.
229. Zuniga-Banuelos, F.J., G. Lemke, M. Hoffmann, U. Reichl, and E. Rapp, *Immunoglobulin A carries sulfated and O-acetylated N-glycans primarily at the tailpiece site – strategies for site-specific N-glycan identification*. Front. Mol. Biosci., 2025. **12**. 1595173.DOI: 10.3389/fmolb.2025.1595173.
230. Schlosser, A. and R. Volkmer-Engert, *Volatile polydimethylcyclsiloxanes in the ambient laboratory air identified as source of extreme background signals in nanoelectrospray mass spectrometry*. J. Mass Spectrom., 2003. **38**(5): p. 523-525.DOI: 10.1002/jms.465.
231. Olsen, J.V., L.M. de Godoy, G. Li, B. Macek, et al., *Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap*. Mol. Cell Proteomics, 2005. **4**(12): p. 2010-2021.DOI: 10.1074/mcp.T500030-MCP200.
232. UniProt, C., *UniProt: the Universal Protein Knowledgebase in 2023*. Nucleic Acids Res, 2023. **51**(D1): p. D523-D531.DOI: 10.1093/nar/gkac1052.
233. Cooper, C.A., E. Gasteiger, and N.H. Packer, *GlycoMod--a software tool for determining glycosylation compositions from mass spectrometric data*. Proteomics, 2001. **1**(2): p. 340-349.DOI: 10.1002/1615-9861(200102)1:2<340::AID-PROT340>3.0.CO;2-B.
234. Uhlen, M., M.J. Karlsson, A. Hober, A.S. Svensson, et al., *The human secretome*. Sci. Signal., 2019. **12**(609).DOI: 10.1126/scisignal.aaz0274.
235. Protein Atlas, C., *Human Protein Atlas version 21.1. The human secretome: secreted to blood*. 2022.
236. UniProt, C. *Customized table showing glycosylated proteins in the reviewed human proteome*. 2022 01.04.2024 [cited 2022 18.10.2022]; Available from: https://www.uniprot.org/uniprotkb?facets=reviewed%3Atrue%2Cmodel_organism%3A9606&fields=accession%2Creviewed%2Cid%2Cprotein_name%2Corganism_name%2Cft_carbohydr&query=human&view=table.
237. UniProt, C., *UniProt: the universal protein knowledgebase in 2021*. Nucleic. Acids Res., 2021. **49**(D1): p. D480-D489.DOI: 10.1093/nar/gkaa1100.
238. Wang, W., *Glycomedicine: The Current State of the Art*. Engineering, 2023. **26**: p. 12-15.DOI: 10.1016/j.eng.2022.03.009.
239. Bond, A., A. Alavi, J.S. Axford, B.E. Bourke, et al., *A detailed lectin analysis of IgG glycosylation, demonstrating disease specific changes in terminal galactose and N-acetylglucosamine*. J. Autoimmun., 1997. **10**(1): p. 77-85.DOI: 10.1006/jaut.1996.0104.
240. Zabczynska, M., P. Link-Lenczowski, M. Novokmet, T. Martin, et al., *Altered N-glycan profile of IgG-depleted serum proteins in Hashimoto's thyroiditis*. Biochim. Biophys. Acta. Gen. Subj., 2020. **1864**(3).DOI: 10.1016/j.bbagen.2019.129464.
241. Lacki, J.K., W. Porawska, U. Mackiewicz, S. Mackiewicz, and W. Muller, *Changes in agalactosyl IgG levels correlate with radiological progression in early rheumatoid arthritis*. Ann. Med., 1996. **28**(3): p. 265-269.DOI: 10.3109/07853899609033129.
242. Pleass, R.J., *The therapeutic potential of sialylated Fc domains of human IgG*. Mabs, 2021. **13**(1). 1953220.DOI: 10.1080/19420862.2021.1953220.
243. Liu, M.Q., W.F. Zeng, P. Fang, W.Q. Cao, 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**(438).DOI: 10.1038/s41467-017-00535-2.

244. Cao, Q.C., X.Y. Zhao, Q. Zhao, X.D. Lv, et al., *Strategy Integrating Stepped Fragmentation and Glycan Diagnostic Ion-Based Spectrum Refinement for the Identification of Core Fucosylated Glycoproteome Using Mass Spectrometry*. *Anal. Chem.*, 2014. **86**(14): p. 6804-6811.DOI: 10.1021/ac501154a.
245. She, Y.M., X. Li, and T.D. Cyr, *Remarkable Structural Diversity of N-Glycan Sulfation on Influenza Vaccines*. *Anal. Chem.*, 2019. **91**(8): p. 5083-5090.DOI: 10.1021/acs.analchem.8b05372.
246. Bochner, B.S., *Siglec-8 on human eosinophils and mast cells, and Siglec-F on murine eosinophils, are functionally related inhibitory receptors*. *Clin. Exp. Allergy.*, 2009. **39**(3): p. 317-324.DOI: 10.1111/j.1365-2222.2008.03173.x.
247. Bochner, B.S., R.A. Alvarez, P. Mehta, N.V. Bovin, et al., *Glycan array screening reveals a candidate ligand for Siglec-8*. *J. Biol. Chem.*, 2005. **280**(6): p. 4307-12.DOI: 10.1074/jbc.M412378200.
248. Sleat, D.E., Y.H. Wang, I. Sohar, H. Lackland, et al., *Identification and validation of mannose 6-phosphate glycoproteins in human plasma reveal a wide range of lysosomal and non-lysosomal proteins*. *Mol. Cell. Proteom.*, 2006. **5**(10): p. 1942-1956.DOI: 10.1074/mcp.M600030-MCP200.
249. Leksa, V., A. Ilkova, K. Vicikova, and H. Stockinger, *Unravelling novel functions of the endosomal transporter mannose 6-phosphate/insulin-like growth factor receptor (CD222) in health and disease: An emerging regulator of the immune system*. *Immunol. Lett.*, 2017. **190**: p. 194-200.DOI: 10.1016/j.imlet.2017.08.011.
250. Budimir, I. *visProteomics*. 2021 [cited 2021 03.11.2021]; Available from: <https://github.com/imforfuture/visProteomics>.
251. Helm, J., L. Hirtler, and F. Altmann, *Towards Mapping of the Human Brain N-Glycome with Standardized Graphitic Carbon Chromatography*. *Biomolecules*, 2022. **12**(1). 85.DOI: 10.3390/biom12010085.
252. Caval, T., J. Zhu, W. Tian, S. Remmelzwaal, et al., *Targeted Analysis of Lysosomal Directed Proteins and Their Sites of Mannose-6-phosphate Modification*. *Mol. Cell. Proteomics*, 2019. **18**(1): p. 16-27.DOI: 10.1074/mcp.RA118.000967.
253. Wuhrer, M., C.A. Koeleman, C.H. Hokke, and A.M. Deelder, *Mass spectrometry of proton adducts of fucosylated N-glycans: fucose transfer between antennae gives rise to misleading fragments*. *Rapid Commun. Mass Spectrom.*, 2006. **20**(11): p. 1747-1754.DOI: 10.1002/rcm.2509.
254. Protein Atlas, C. *CRISPLD1-Protein detected in human plasma by mass spectrometry*. 2022 31.05.2022 26.10.2022]; Available from: [v21.proteinatlas.org/ENSG00000121005-CRISPLD1/blood+protein](https://www.proteinatlas.org/ENSG00000121005-CRISPLD1/blood+protein).
255. Farrah, T., E.W. Deutsch, G.S. Omenn, D.S. Campbell, et al., *A high-confidence human plasma proteome reference set with estimated concentrations in PeptideAtlas*. *Mol. Cell. Proteomics.*, 2011. **10**(9): p. 1-14.DOI: 10.1074/mcp.M110.006353.
256. National Center for Biotechnology Information. *PubChem Compound Summary for CID 439197, N-Acetyl-Neuraminic Acid*. 2004 April 16,2023 [cited 2022 13.04.2022]; September 16, 2004:[Available from: <https://pubchem.ncbi.nlm.nih.gov/compound/N-Acetyl-Neuraminic-Acid>].
257. Wang, F., J. Liigand, S. Tian, D. Arndt, et al., *CFM-ID 4.0: More Accurate ESI-MS/MS Spectral Prediction and Compound Identification*. *Anal. Chem.*, 2021. **93**(34): p. 11692-11700.DOI: 10.1021/acs.analchem.1c01465.
258. National Center for Biotechnology Information. *PubChem Compound Summary for CID 59861591*. 2012 April 16, 2023 [cited 2022 13.04.2022]; August 20, 2012:[Available from: <https://pubchem.ncbi.nlm.nih.gov/compound/59861591>].
259. National Center for Biotechnology Information. *PubChem Compound Summary for CID 59173067*. 2012 April 16, 2023 [cited 2022 13.04.2022]; August 20, 2012:[Available from: <https://pubchem.ncbi.nlm.nih.gov/compound/59173067>].
260. Ihlenfeldt, W.D., E.E. Bolton, and S.H. Bryant, *The PubChem chemical structure sketcher*. *J. Cheminform.*, 2009. **1**(20). 20.DOI: 10.1186/1758-2946-1-20.
261. Campbell, M.P., T. Nguyen-Khuong, C.A. Hayes, S.A. Flowers, et al., *Validation of the curation pipeline of UniCarb-DB: building a global glycan reference MS/MS repository*. *Biochim. Biophys. Acta*, 2014. **1844**(1 Pt A): p. 108-116.DOI: 10.1016/j.bbapap.2013.04.018.
262. Alocci, D., J. Mariethoz, A. Gastaldello, E. Gasteiger, et al., *GlyConnect: Glycoproteomics Goes Visual, Interactive, and Analytical*. *J. Proteome. Res.*, 2019. **18**(2): p. 664-677.DOI: 10.1021/acs.jproteome.8b00766.
263. Yamada, I., M. Shiota, D. Shinmachi, T. Ono, et al., *The GlyCosmos Portal: a unified and comprehensive web resource for the glycosciences*. *Nat. Methods*, 2020. **17**(7): p. 649-650.DOI: 10.1038/s41592-020-0879-8.
264. Expasy. *GlyConnect*. 2019 [cited 2022 01.08.2022]; Available from: <https://glyconnect.expasy.org/>.

265. van Rooijen, J.J., J.P. Kamerling, and J.F. Vliegthart, *Sulfated di-, tri- and tetraantennary N-glycans in human Tamm-Horsfall glycoprotein*. Eur. J. Biochem., 1998. **256**(2): p. 471-487.DOI: 10.1046/j.1432-1327.1998.2560471.x.
266. Lippold, S., A. Buttner, M.S.F. Choo, M. Hook, et al., *Cysteine Aminoethylation Enables the Site-Specific Glycosylation Analysis of Recombinant Human Erythropoietin using Trypsin*. Anal. Chem., 2020. **92**(14): p. 9476-9481.DOI: 10.1021/acs.analchem.0c01794.
267. Ohta, M., T. Ohnishi, Y.A. Ioannou, M.E. Hodgson, et al., *Human alpha-N-acetylgalactosaminidase: site occupancy and structure of N-linked oligosaccharides*. Glycobiology, 2000. **10**(3): p. 251-261.DOI: 10.1093/glycob/10.3.251.
268. Wessels, H.J.C.T., P. Kulkarni, M. van Dael, A. Suppers, et al., *Plasma glycoproteomics delivers high-specificity disease biomarkers by detecting site-specific glycosylation abnormalities*. BioRxiv, 2022.DOI: 10.1101/2022.05.31.494121.
269. Kreimer, S., M.E. Belov, W.F. Danielson, L.I. Levitsky, et al., *Advanced Precursor Ion Selection Algorithms for Increased Depth of Bottom-Up Proteomic Profiling*. J. Proteome Res., 2016. **15**(10): p. 3563-3573.DOI: 10.1021/acs.jproteome.6b00312.
270. Wuhrer, M., C.A.M. Koeleman, and A.M. Deelder, *Hexose Rearrangements upon Fragmentation of N-Glycopeptides and Reductively Aminated N-Glycans*. Anal. Chem., 2009. **81**(11): p. 4422-4432.DOI: 10.1021/ac900278q.
271. Acs, A., O. Ozohanics, K. Vekey, L. Drahos, and L. Turiak, *Distinguishing Core and Antenna Fucosylated Glycopeptides Based on Low-Energy Tandem Mass Spectra*. Anal. Chem., 2018. **90**(21): p. 12776-12782.DOI: 10.1021/acs.analchem.8b03140.
272. Vreeker, G.C. and M. Wuhrer, *Reversed-phase separation methods for glycan analysis*. Anal. Bioanal. Chem., 2017. **409**(2): p. 359-378.DOI: 10.1007/s00216-016-0073-0.
273. Ji, E.S., H.K. Lee, G.W. Park, K.H. Kim, et al., *Isomer separation of sialylated O- and N-linked glycopeptides using reversed-phase LC-MS/MS at high temperature*. J. Chromatogr. B. Analyt. Technol. Biomed. Life Sci., 2019. **1110-1111**: p. 101-107.DOI: 10.1016/j.jchromb.2019.02.015.
274. Cumming, D.A., C.G. Hellerqvist, M. Harris-Brandts, S.W. Michnick, et al., *Structures of asparagine-linked oligosaccharides of the glycoprotein fetuin having sialic acid linked to N-acetylglucosamine*. Biochemistry, 1989. **28**(15): p. 6500-6512.DOI: 10.1021/bi00441a051.
275. Guttman, M. and K.K. Lee, *Site-Specific Mapping of Sialic Acid Linkage Isomers by Ion Mobility Spectrometry*. Anal. Chem., 2016. **88**(10): p. 5212-5217.DOI: 10.1021/acs.analchem.6b00265.
276. Hinneburg, H., J. Hofmann, W.B. Struwe, A. Thader, et al., *Distinguishing N-acetylneuraminic acid linkage isomers on glycopeptides by ion mobility-mass spectrometry*. Chem. Commun (Camb), 2016. **52**(23): p. 4381-4384.DOI: 10.1039/c6cc01114d.
277. Pett, C., W. Nasir, C. Sihlbom, B.M. Olsson, et al., *Effective Assignment of alpha2,3/alpha2,6-Sialic Acid Isomers by LC-MS/MS-Based Glycoproteomics*. Angew. Chem. Int. Ed. Engl., 2018. **57**(30): p. 9320-9324.DOI: 10.1002/anie.201803540.
278. Chen, C.H., Y.P. Lin, J.L. Lin, S.T. Li, et al., *Rapid Identification of Terminal Sialic Acid Linkage Isomers by Pseudo-MS Mass Spectrometry*. Isr. J. Chem., 2015. **55**(3-4): p. 412-422.DOI: 10.1002/ijch.201400141.
279. Tsuchida, A., T. Okajima, K. Furukawa, T. Ando, et al., *Synthesis of disialyl Lewis a (Le(a)) structure in colon cancer cell lines by a sialyltransferase, ST6GalNAc VI, responsible for the synthesis of alpha-series gangliosides*. J. Biol. Chem., 2003. **278**(25): p. 22787-22794.DOI: 10.1074/jbc.M211034200.
280. Miyazaki, K., K. Ohmori, M. Izawa, T. Koike, et al., *Loss of disialyl Lewis(a), the ligand for lymphocyte inhibitory receptor sialic acid-binding immunoglobulin-like lectin-7 (Siglec-7) associated with increased sialyl Lewis(a) expression on human colon cancers*. Cancer Res., 2004. **64**(13): p. 4498-4505.DOI: 10.1158/0008-5472.CAN-03-3614.
281. Campos, D., M. Girgis, Q. Yang, G. Zong, et al., *"Ghost" Fragment Ions in Structure and Site-Specific Glycoproteomics Analysis*. Anal. Chem., 2023. **95**(27): p. 10145-10148.DOI: 10.1021/acs.analchem.3c02207.
282. Yamada, K., H. Kayahara, M. Kinoshita, and S. Suzuki, *Simultaneous Analysis of Sulfated and Phosphorylated Glycans by Serotonin-Immobilized Column Enrichment and Hydrophilic Interaction Chromatography*. Anal. Chem., 2018. **90**(14): p. 8387-8395.DOI: 10.1021/acs.analchem.8b00714.

283. Zhang, Y., E.P. Go, H. Jiang, and H. Desaire, *A novel mass spectrometric method to distinguish isobaric monosaccharides that are phosphorylated or sulfated using ion-pairing reagents*. J. Am. Soc. Mass Spectrom., 2005. **16**(11): p. 1827-1839.DOI: 10.1016/j.jasms.2005.07.010.
284. Protein Atlas, C., *Normal tissue data available from Human Protein Atlas*. 2019, The Human Protein Atlas.
285. Bochner, B.S., R.A. Alvarez, P. Mehta, N.V. Bovin, et al., *Glycan array screening reveals a candidate ligand for Siglec-8*. J. Biol. Chem., 2005. **280**(6): p. 4307-4312.DOI: 10.1074/jbc.M412378200.
286. Klaric, T.S. and G. Lauc, *The dynamic brain N-glycome*. Glycoconj. J., 2022. **39**: p. 443-471.DOI: 10.1007/s10719-022-10055-x.
287. Castillo, G., R. Kleene, M. Schachner, G. Loers, and A.E. Torda, *Proteins Binding to the Carbohydrate HNK-1: Common Origins?* Int. J. Mol. Sci., 2021. **22**(15). 8116.DOI: 10.3390/ijms22158116.
288. Wang, W., T. Zhang, J. Nouta, P.A. van Veelen, et al., *Human Prostate-Specific Antigen Carries N-Glycans with Ketodeoxynononic Acid*. Engineering, 2023. **26**: p. 119-131.DOI: 10.1016/j.eng.2023.02.009.
289. Lämmle, B. and J.H. Griffin, *Prothrombin and Fibrin clotting factor: the balance of procoagulant and inhibitory factors*. Clin. Haematol., 1985. **14**(2): p. 281-342.
290. Gomes, M.M., S.B. Wall, K. Takahashi, J. Novak, et al., *Analysis of IgA1 N-glycosylation and its contribution to FcαRI binding*. Biochemistry, 2008. **47**(43): p. 11285-11299.DOI: 10.1021/bi801185b.
291. Mattu, T.S., R.J. Pleass, A.C. Willis, M. Kilian, et al., *The glycosylation and structure of human serum IgA1, Fab, and Fc regions and the role of N-glycosylation on FcαRI receptor interactions*. J. Biol. Chem., 1998. **273**(4): p. 2260-2272.DOI: 10.1074/jbc.273.4.2260.
292. Goonatilake, E., J.T. Smilowitz, K.V. Marino, B.J. German, et al., *Immunoglobulin A N-glycosylation Presents Important Body Fluid-specific Variations in Lactating Mothers*. Mol. Cell. Proteom., 2019. **18**(11): p. 2165-2177.DOI: 10.1074/mcp.RA119.001648.
293. Liao, H.Y., C.H. Hsu, S.C. Wang, C.H. Liang, et al., *Differential receptor binding affinities of influenza hemagglutinins on glycan arrays*. J. Am. Chem. Soc., 2010. **132**(42): p. 14849-14856.DOI: 10.1021/ja104657b.
294. Wang, C.C., J.R. Chen, Y.C. Tseng, C.H. Hsu, et al., *Glycans on influenza hemagglutinin affect receptor binding and immune response*. Proc. Natl. Acad. Sci. U. S. A., 2009. **106**(43): p. 18137-18142.DOI: 10.1073/pnas.0909696106.
295. Lo-Guidice, J.M., J.M. Wieruszkeski, J. Lemoine, A. Verbert, et al., *Sialylation and sulfation of the carbohydrate chains in respiratory mucins from a patient with cystic fibrosis*. J. Biol. Chem., 1994. **269**(29): p. 18794-18813.
296. Mowery, P., Z.Q. Yang, E.J. Gordon, O. Dwir, et al., *Synthetic glycoprotein mimics inhibit L-selectin-mediated rolling and promote L-selectin shedding*. Chem. Biol., 2004. **11**(5): p. 725-732.DOI: 10.1016/j.chembiol.2004.03.027.
297. Otsuki, S., S.R. Hanson, S. Miyaki, S.P. Grogan, et al., *Extracellular sulfatases support cartilage homeostasis by regulating BMP and FGF signaling pathways*. Proc Natl Acad Sci U S A, 2010. **107**(22): p. 10202-7.DOI: 10.1073/pnas.0913897107.
298. Mayboroda, O.A., G.S.M. Lageveen-Kammeijer, M. Wuhrer, and R. Dolhain, *An Integrated Glycosylation Signature of Rheumatoid Arthritis*. Biomolecules, 2023. **13**(7). 1106.DOI: 10.3390/biom13071106.
299. Xu, Z., Y. Liu, S. He, R. Sun, et al., *Integrative Proteomics and N-Glycoproteomics Analyses of Rheumatoid Arthritis Synovium Reveal Immune-Associated Glycopeptides*. Mol. Cell. Proteomics, 2023. **22**(5). 100540.DOI: 10.1016/j.mcpro.2023.100540.
300. Dotz, V., A. Visconti, H.J. Lomax-Browne, F. Clerc, et al., *O- and N-Glycosylation of Serum Immunoglobulin A is Associated with IgA Nephropathy and Glomerular Function*. J. Am. Soc. Nephrol., 2021. **32**(10): p. 2455-2465.DOI: 10.1681/Asn.2020081208.
301. Klapoetke, S.C., J. Zhang, and S. Becht, *Glycosylation characterization of Human IgA1 with differential deglycosylation by UPLC-ESI TOF MS*. J. Pharm. Biomed. Anal., 2011. **56**(3): p. 513-520.DOI: 10.1016/j.jpba.2011.06.010.
302. Atkin, J.D., R.J. Pleass, R.J. Owens, and J.M. Woof, *Mutagenesis of the human IgA1 heavy chain tailpiece that prevents dimer assembly*. Journal of Immunology, 1996. **157**(1): p. 156-159.
303. Bastian, A., H. Kratzin, K. Eckart, and N. Hilschmann, *Intra- and interchain disulfide bridges of the human J chain in secretory immunoglobulin A*. Biol Chem Hoppe Seyler, 1992. **373**(12): p. 1255-63.DOI: 10.1515/bchm3.1992.373.2.1255.

304. Vaerman, J.P., K. Hagiwara, K. Kobayashi, and M. Rits, *Complexes of albumin and alpha 1-antitrypsin with Fc-fragment of IgA monomer are disulfide-bound to penultimate C-terminal cysteine in the C alpha 3-domain*. Immunol. Lett., 1987. **15**(1): p. 67-72.DOI: 10.1016/0165-2478(87)90078-2.
305. Reinhart, D. and R. Kunert, *Upstream and downstream processing of recombinant IgA*. Biotechnol. Lett., 2015. **37**(2): p. 241-251.DOI: 10.1007/s10529-014-1686-z.
306. Sun, W., Q. Zhang, X. Zhang, N.H. Tran, et al., *Glycopeptide database search and de novo sequencing with PEAKS GlycanFinder enable highly sensitive glycoproteomics*. Nat. Commun., 2023. **14**(1). 4046.DOI: 10.1038/s41467-023-39699-5.
307. Doi, T., K. Kanatsu, K. Sekita, H. Yoshida, et al., *Detection of IgA class circulating immune complexes bound to anti-C3d antibody in patients with IgA nephropathy*. J. Immunol. Methods, 1984. **69**(1): p. 95-104.DOI: 10.1016/0022-1759(84)90281-3.
308. Martinez-Flores, J.A., M. Serrano, D. Perez, D. Lora, et al., *Detection of circulating immune complexes of human IgA and beta 2 glycoprotein I in patients with antiphospholipid syndrome symptomatology*. J. Immunol. Methods, 2015. **422**: p. 51-58.DOI: 10.1016/j.jim.2015.04.002.
309. Hemmerich, S., C.R. Bertozzi, H. Leffler, and S.D. Rosen, *Identification of the sulfated monosaccharides of GlyCAM-1, an endothelial-derived ligand for L-selectin*. Biochemistry, 1994. **33**(16): p. 4820-4829.DOI: 10.1021/bi00182a010.
310. Wang, Y.F., C.F. Chang, C.Y. Chi, H.C. Wang, et al., *Characterization of glycan binding specificities of influenza B viruses with correlation with hemagglutinin genotypes and clinical features*. J. Med. Virol., 2012. **84**(4): p. 679-685.DOI: 10.1002/jmv.23219.
311. Scharfman, A., P. Delmotte, J. Beau, G. Lamblin, et al., *Sialyl-Le(x) and sulfo-sialyl-Le(x) determinants are receptors for P. aeruginosa*. Glycoconj. J., 2000. **17**(10): p. 735-740.DOI: 10.1023/a:1011091112884.
312. Rogers, G.N., G. Herrler, J.C. Paulson, and H.D. Klenk, *Influenza-C Virus Uses 9-O-Acetyl-N-Acetylneuraminic Acid as a High-Affinity Receptor Determinant for Attachment to Cells*. Journal of Biological Chemistry, 1986. **261**(13): p. 5947-5951.
313. Sjöberg, E.R., L.D. Powell, A. Klein, and A. Varki, *Natural Ligands of the B-Cell Adhesion Molecule Cd22-Beta Can Be Masked by 9-O-Acetylation of Sialic Acids*. Journal of Cell Biology, 1994. **126**(2): p. 549-562.DOI: DOI 10.1083/jcb.126.2.549.
314. Weiman, S., S. Dahesh, A.F. Carlin, A. Varki, et al., *Genetic and biochemical modulation of sialic acid O-acetylation on group B Streptococcus: phenotypic and functional impact*. Glycobiology, 2009. **19**(11): p. 1204-1213.DOI: 10.1093/glycob/cwp111.
315. Li, H., A.W.T. Chiang, and N.E. Lewis, *Artificial intelligence in the analysis of glycosylation data*. Biotechnol. Adv., 2022. **60**(108008): p. 108008.DOI: 10.1016/j.biotechadv.2022.108008.

List of Figures

| | |
|---|----|
| Figure 2.1. Diversity of glycoconjugates in animal cells..... | 5 |
| Figure 2.2. Protein N-glycosylation..... | 8 |
| Figure 2.3. N-glycan processing pathway. | 9 |
| Figure 2.4. Structural complexity of N-glycans. | 13 |
| Figure 2.5. Methodologies for the analysis of glycoproteins. | 15 |
| Figure 2.6. Distribution of blood plasma proteins over a concentration range of more than ten orders of magnitude. | 17 |
| Figure 2.7. Fragmentation of a complex-type N-glycopeptide..... | 23 |
| Figure 2.8. IgA functions in the immune system influenced by N-glycosylation..... | 28 |
| Figure 2.9. Comparison of incidence rate in 1999 versus 2019 for Alzheimer's disease and other dementias as cause of death or injury. | 29 |
| Figure 2.10. Comparison of incidence rate in 1999 versus 2019 for pancreatic cancer as a cause of death or injury. | 30 |
| Figure 2.11. Comparison of incidence rate in 1999 versus 2019 for rheumatoid arthritis as cause of death or injury. | 31 |
| Figure 2.12. Examples about N-glycans displayed in human serum IgG in different AIDs. | 32 |
| Figure 3.1. Sample preparation workflow and measurement..... | 37 |
| Figure 3.2. Human blood plasma N-glycoproteomic data analysis workflow. | 43 |
| Figure 3.3. Experimental design for the site-specific identification of sulfated and other rare N-glycans in IgA..... | 46 |
| Figure 3.4. Expanding the N-glycoproteomic analysis of human serum IgA by applying multiple strategies. | 49 |
| Figure 4.1. Exploring the low-abundant blood plasma glycoproteome. | 53 |
| Figure 4.2. Gain achieved by applying the established sample preparation workflow..... | 57 |
| Figure 4.3. Validation of N-glycopeptide identifications. | 61 |
| Figure 4.4. Evaluation of the evidence present in fucose-containing gPSMs..... | 64 |
| Figure 4.5. Manual revision of an incorrect identification for β -2-glycoprotein 1 (peptide K.LGnWSAMPSC.A) containing a difucosylated complex-type N-glycan. | 66 |
| Figure 4.6. Annotation of fragment ions released from sulfated N-glycopeptide. | 68 |
| Figure 4.7. Annotation of fragment ions released from phosphorylated N-glycopeptide..... | 69 |

| | |
|--|-----|
| Figure 4.8. Annotation of fragment ions released from typical and rare N-glycans attached to N121 (peptide GHVNITR). | 71 |
| Figure 4.9. N-Glycan de novo sequencing of N-glycopeptides with disialo-antennary N-glycan structure confirming HexNAc-NeuAc linkage..... | 74 |
| Figure 5.1. Description of diverse sulfated N-glycans identified by us at the N-glycosylation sites of human serum IgA1 and IgA2..... | 85 |
| Figure 5.2. Comparison of HCD.step fragment ion spectra derived from two N-glycopeptides (with the same peptide backbone)—one bearing a sulfated and one a non-sulfated N-glycan at the N-glycosylation site in the tailpiece. | 92 |
| Figure 5.3. Site-specific description of the here identified N-glycans carrying sulfated HexNAc and O-acetylated NeuAc in human serum IgA. | 95 |
| Figure 5.4. Comparison of micro-heterogeneity of N-glycosylation identified in commercially available human serum IgA samples from two suppliers. | 100 |
| Figure 5.5. A comparative analysis of sulfated N-glycans with and without HexNAc-sulfated oxonium ions attached to the IgA tailpiece site in commercially available human serum IgA samples from two suppliers. | 101 |
| Figure 6.1. Gain obtained with the workflow developed for the in depth-N-glycoproteomic analysis of human blood plasma proteins..... | 107 |

List of Supplementary Figures

A.1. In-depth analysis of the low abundant *N*-glycoproteome of human blood plasma

| | |
|--|-----|
| Supplementary Figure A.1.1. Oxonium ion filter applied to identify oxonium ions present in sulfated glycopeptides..... | 137 |
| Supplementary Figure A.1.2. Oxonium ion filters applied to identify oxonium ions present in phosphorylated glycopeptides. | 137 |
| Supplementary Figure A.1.3. Comparison of the gain of the new workflow with respect to the identification of glycoprotein, N-glycopeptide and N-glycan compositions. | 138 |
| Supplementary Figure A.1.4. N-glycopeptides identified in blood plasma after in-depth N-glycoproteomic analysis. | 139 |
| Supplementary Figure A.1.5. Example of a gPSM with the features assigned for the validation category named “True Match with Outstanding Evidence.” | 140 |

| | |
|---|-----|
| Supplementary Figure A.1.6. Example of a gPSM with the features assigned for the validation category named “True Match with Evidence.” | 141 |
| Supplementary Figure A.1.7. Example of a gPSM with the features assigned for the validation category named “True Match with Evidence and Alternatives.” | 142 |
| Supplementary Figure A.1.8. Example of a gPSM with the features assigned for the validation category named as “True Match with Evidence Related to matches from the same peptide.” | 143 |
| Supplementary Figure A.1.9. Example of a gPSM with the features assigned for the validation category named “Quasi-true Match Double Site.” | 144 |
| Supplementary Figure A.1.10. Example of a gPSM with the features assigned for the validation category named “Quasi-true Match change glycan.” | 145 |
| Supplementary Figure A.1.11. Example of a gPSM with the features assigned for the validation category named “Quasi-true corrected Match.” | 146 |
| Supplementary Figure A.1.12. Example of a gPSM with the features assigned for the validation category named “Uncertain-Change glycopeptide match.” | 147 |
| Supplementary Figure A.1.13. Example of a gPSM with the features assigned for the validation category named “Uncertain glycopeptide.” | 148 |
| Supplementary Figure A.1.14. Example of a gPSM with the features assigned for the validation category named “False O-glycan.” | 149 |
| Supplementary Figure A.1.15. Example of a gPSM with the features assigned for the validation category named “False.” | 150 |
| Supplementary Figure A.1.16. Examples of gPSM including multiple fucose in a complex N-glycan with a potential error in the assignation of the monoisotopic peak. | 151 |
| Supplementary Figure A.1.17. Additional examples of gPSM including multiple fucose in a complex N-glycans with a potential error assigning the monoisotopic peak..... | 152 |
| Supplementary Figure A.1.18. Byonic™ software annotation of fragment ions released from sulfated N-glycopeptide. | 153 |
| Supplementary Figure A.1.19. Byonic™ software annotation of fragment ions released from phosphorylated N-glycopeptide..... | 154 |
| Supplementary Figure A.1.20. Searching for rare N-glycopeptides in untreated blood plasma and top 14-HAP depleted samples. | 155 |
| Supplementary Figure A.1.21. Byonic™ software annotation of HCD.step fragment ion spectra containing rare N-glycans found in N121 from prothrombin (UniProtKB P00734). | 156 |
| Supplementary Figure A.1.22. ESI-MS/MS prediction for two molecules proposed as the N-glycan building blocks of the rare N-glycopeptide identifications observed. | 157 |

| | |
|--|-----|
| Supplementary Figure A.1.23. Byonic™ software annotation of HCD.step fragment ion spectra from N-glycopeptides with disialo-antennary structure confirming HexNAc-NeuAc linkage. | 158 |
|--|-----|

A.2. Site-specific identification of sulfated and *O*-acetylated *N*-glycans in human serum IgA

| | |
|---|-----|
| Supplementary Figure A.2.1. Separation of primarily light and heavy chains of human serum IgA subclasses by GELFrEE® system..... | 159 |
| Supplementary Figure A.2.2. Annotated data of a glycopeptide bearing the sulfated N-glycan HexNAc(4)Hex(5)Fuc(1)NeuAc(2)Sulfo(1). | 160 |
| Supplementary Figure A.2.3. Annotated data of a glycopeptide bearing the sulfated N-glycan HexNAc(4)Hex(5)NeuAc(1)Sulfo(1). | 161 |
| Supplementary Figure A.2.4. Annotated data of a glycopeptide bearing the sulfated N-glycan: HexNAc(4)Hex(5)NeuAc(2)Sulfo(1). | 162 |
| Supplementary Figure A.2.5. Annotated spectra of a glycopeptide bearing the sulfated N-glycan: HexNAc(4)Hex(5)Fuc(1)NeuAc(2)Sulfo(2). | 163 |
| Supplementary Figure A.2.6. Annotated data of a glycopeptide bearing the sulfated N-glycan: HexNAc(5)Hex(5)Fuc(1)NeuAc(2)Sulfo(1). | 164 |
| Supplementary Figure A.2.7. Annotated spectra of a glycopeptide bearing the sulfated N-glycan: HexNAc(4)Hex(6)Fuc(2)NeuAc(1)Sulfo(1). | 165 |
| Supplementary Figure A.2.8. Annotated spectra of a glycopeptide bearing the sulfated N-glycan: HexNAc(4)Hex(5)Fuc(1)NeuAc(1)Sulfo(1). | 166 |
| Supplementary Figure A.2.9. Annotated data of a glycopeptide containing NeuAc <i>O</i> -acetylation in the N-glycan HexNAc(4)Hex(5)Fuc(1)NeuAc(2)Ac(1)..... | 167 |
| Supplementary Figure A.2.10. Annotated spectra of a glycopeptide containing the NeuAc <i>O</i> -acetylated N-glycan: HexNAc(5)Hex(5)Fuc(1)NeuAc(1)Ac(1)..... | 168 |
| Supplementary Figure A.2.11. Annotated data of glycopeptide containing NeuAc <i>O</i> -acetylation in the N-glycan HexNAc(5)Hex(5)Fuc(1)NeuAc(2)Ac(1)..... | 169 |
| Supplementary Figure A.2.12. Annotated data of a glycopeptide containing NeuAc <i>O</i> -acetylation plus HexNAc sulfation in the N-glycan: HexNAc(4)Hex(5)Fuc(1)NeuAc(2)Ac(1)Sulfo(1). | 170 |
| Supplementary Figure A.2.13. Annotated data of a glycopeptide potentially containing HexA sulfation in the N-glycan HexNAc(4)Hex(5)NeuAc(1)HexA(1)Sulfo(1) of a contaminant protein. | 171 |
| Supplementary Figure A.2.14. Annotated data of a glycopeptide potentially containing HexA in the N-glycan HexNAc(4)Hex(5)NeuAc(1)HexA(1) of a contaminant protein..... | 172 |

List of Tables

| | |
|---|----|
| Table 2.1. Typical N-glycan building blocks in humans..... | 7 |
| Table 4.1. Micro-heterogeneity of N-glycosylation observed in zinc- α -2-glycoprotein (P25311) after each step of the sample preparation workflow. | 58 |
| Table 4.2. N-Glycan-derived marker ions. | 63 |
| Table 5.1. Cross-referenced N-glycans bearing a sulfated-sugar and O-acetylated NeuAc. | 94 |
| Table 5.2. Relative abundance of IgA subclasses of each human serum IgA commercial sample. | 97 |

List of Supplementary Tables

All tables are found only digital in Excel files uploaded to MPI server and to the cloud:

<https://nextcloud.mpi-magdeburg.mpg.de/index.php/s/X8rsEPQRAAHY7Zi>

A.1. In-depth analysis of the low abundant *N*-glycoproteome of human blood plasma

Supplementary Table A.1.1. Proteomic analysis of HILIC-fractions from untreated blood plasma sample.

Supplementary Table A.1.2. Proteomic analysis of HILIC-fractions of top 14-HAP depleted sample.

Supplementary Table A.1.3. Proteomic analysis of HILIC-fractions of top 14-HAP depleted sample plus fractionation by protein molecular weight (applying the developed workflow).

Supplementary Table A.1.4. *N*-glycoproteomic analysis of untreated blood plasma sample.

Supplementary Table A.1.5. *N*-glycoproteomic analysis of top 14-HAP depleted sample.

Supplementary Table A.1.6. HCD.step-HPB list. *N*-glycoproteomic analysis of top 14-HAP depleted sample plus fractionation by protein molecular weight.

Supplementary Table A.1.7. HCD.low-HPB list. *N*-glycoproteomic analysis of top 14-HAP depleted sample plus fractionation by protein molecular weight.

Supplementary Table A.1.8. Classified fucosylated *N*-glycan structures in *N*-glycopeptide identifications validated as true in the HCD.step-HBP list.

Supplementary Table A.1.9. Sulfated *N*-glycopeptide identifications validated true in the HCD.step-HBP list.

Supplementary Table A.1.10. Phosphorylated *N*-glycopeptide identifications validated as true in the HCD.step-HBP list.

Supplementary Table A.1.11. Glucuronidated *N*-glycopeptide identifications validated as true in the HCD.step-HBP list.

Supplementary Table A.1.12. Rare *N*-glycopeptide identifications validated as true in the HCD.step-HBP list.

Supplementary Table A.1.13. Cross-referenced rare compositions and modified *N*-glycans.

Supplementary Table A.1.14. Cross-referenced special *N*-glycan structural features.

A.2. Site-specific identification of sulfated and *O*-acetylated *N*-glycans in human serum IgA

Supplementary Table A.2.1. Identification of MS² spectra containing sulfated HexNAc within GELFrEE® fractionated human serum IgA.

Supplementary Table A.2.2. Search parameters for (*N*-glyco-)proteomic analysis on IgA H-fraction performed through Byonic™.

Supplementary Table A.2.3. Search parameters for (*N*-glyco-)proteomic analysis on the unfractionated human serum IgA samples from supplier 1 and supplier 2 performed through Byonic™.

Supplementary Table A.2.4. (*N*-glyco-)proteomic analysis of IgA H-fraction.

Supplementary Table A.2.5. (*N*-glyco-)proteomic analysis of unfractionated human serum IgA from supplier 1 and supplier 2.

Supplementary Table A.2.6. Proteomic analysis from serum IgA samples from supplier 1 and supplier 2 performed through Proteome Discoverer™.

Supplementary Table A.2.7. Micro-heterogeneity of *N*-glycosylation in ambiguous sites of human IgA subclasses analyzed in the sample from supplier 1.

Supplementary Table A.2.8. Micro-heterogeneity of *N*-glycosylation in ambiguous sites of human IgA subclasses analyzed in the sample from supplier 2.

Supplementary Table A.2.9. Micro-heterogeneity of *N*-glycosylation in unique sites of human IgA2 analyzed in the sample from supplier 1.

Supplementary Table A.2.10. Micro-heterogeneity of *N*-glycosylation in unique sites of human IgA2 analyzed in the sample from supplier 2.

Supplementary Table A.2.11. Dedicated (*N*-glyco-)proteomic analysis of sulfated *N*-glycans (with and without HexNAc-sulfated oxonium ions) in human serum IgA from two suppliers.

Supplementary Table A.2.12. Relative quantification of sulfated *N*-glycans in homologous sites of human IgA subclasses identified with and without HexNAc-sulfated oxonium ions in two commercial human serum IgA samples.

List of Materials and Chemicals

2

2,2,2-trifluoroethanol

#808259, Merck, Darmstadt, Germany

A

Acetonitrile

#A955-212, Fisher Scientific, Schwerte, Germany

Amicon Ultra 15 mL cut-off 3 kDa device

#UFC900396, Merck, County Cork, Ireland

Ammonium bicarbonate

#09830, Merck, Darmstadt, Germany

C

C18 Separation column Acclaim PepMap™ RSLC nanoViper

#164941, Thermo Fisher Scientific, Lithuania

C18 Trap column Acclaim PepMap™ 100

#164199, Thermo Fisher Scientific, Lithuania

Calcium chloride

#A4689, Merck, Darmstadt, Germany

D

DL-dithiothreitol

#D5545, Merck, Darmstadt, Germany

F

Formic acid

#56302, Merck, Darmstadt, Germany

G

GELFrEE® system Cartridge Kit 8%

#42103, Abcam, Cambridge, UK

H

High Select™ Top 14 Abundant Protein Depletion Midi Spin Columns

#A36371, Thermo Fisher Scientific, IL, USA

I

Immunoglobulin A purified from human blood serum

#14036-1MG, Sigma-Aldrich, MO, USA

#16-16-090701, Athens Research & Technology,
Georgia, USA

Iodoacetamide

#11149, Merck, Darmstadt, Germany

N

Nanosep® Omega filters cut-off 10 kDa

#OD010C35, Pall®, Puerto Rico

Novex™ Tris-Glycine Protein Mini Gel 8-16%

#XP08165BOX Thermo Fisher Scientific, CA, USA

P

Pierce™ BCA Protein Assay Kit

#23225, Thermo Fisher Scientific, IL, USA

Pipette Tips RT LTS 250µL SX 768A/8

Mettler Toledo, CA, USA

Potassium chloride

KCl, #104935, Merck, Darmstadt, Germany
Potassium phosphate dibasic
#104873, Merck, Darmstadt, Germany

S

Sodium chloride
#P029.3, Carl Roth, Karlsruhe, Germany
Sodium dodecyl sulfate
#75746, Merck, Darmstadt, Germany
Sodium phosphate dibasic
#106585, Merck, Darmstadt, Germany

T

Trifluoroacetic acid

#28904, Fisher Scientific, Schwerte, Germany
Trypsin sequencing grade
#V5111, Promega, Madison, WI, USA

V

VisuConTM-F Frozen normal control plasma
#FRNCP0105, Affinity Biologicals, ON, Canada

W

Water (Milli-Q LC-MS grade)
Millipore Milli-Q[®] Advantage A10 system equipped with
a LC-Pak[®] Polisher filter unit, Merck Millipore,
Darmstadt, Germany

List of Instruments and Devices

B

Byologic™ v4.4-74-g75311a1df5 x64
Protein Metrics, Cupertino, CA, USA

Byologic™ v4.6-37-gda29558119 x64
Protein Metrics, Cupertino, CA, USA

Byonic™ search engine v4.2.10
Protein Metrics, Cupertino, CA, USA

Byonic™ search engine v4.4.2
Protein Metrics, Cupertino, CA, USA

G

GELFrEE® 8100 Fractionation System
Abcam, Cambridge, UK

H

Heraeus Multifuge X 1R centrifuge
Thermo Fisher Scientific, Osterode, Germany

I

Incubator Titramax 1000
Heidolph, Schwabach, Germany

L

LC-Pak® Polisher filter unit
Merck Millipore, Darmstadt, Germany

M

Mascot version 2.6

Matrix Science, London, UK

Millipore Milli-Q® Advantage A10 system
Merck Millipore, Darmstadt, Germany

MultiBio RS-24 rotor
BIOSAN, Riga, Latvia

O

Orbitrap Eclipse Tribrid Mass Spectrometer
Thermo Fisher Scientific, Dreieich, Germany

Orbitrap Elite Hybrid Ion Trap-Orbitrap Mass Spectrometer
Thermo Fisher Scientific, Bremen, Germany

P

Proteome Discoverer™ version 2.5.0.400
Thermo Fisher Scientific, Bremen, Germany

R

Rotational vacuum concentrator, RVC 2-33 CO plus
Christ, Osterode am Harz, Germany

T

Thermo Xcalibur Qual Browser version 2.2
Thermo Fisher Scientific, Bremen, Germany

ThermoMixer C
Eppendorf, Hamburg, Germany

U

UHPLC, Dionex UltiMate 3000 RSLCnano system
Thermo Fisher Scientific, Germering, Germany

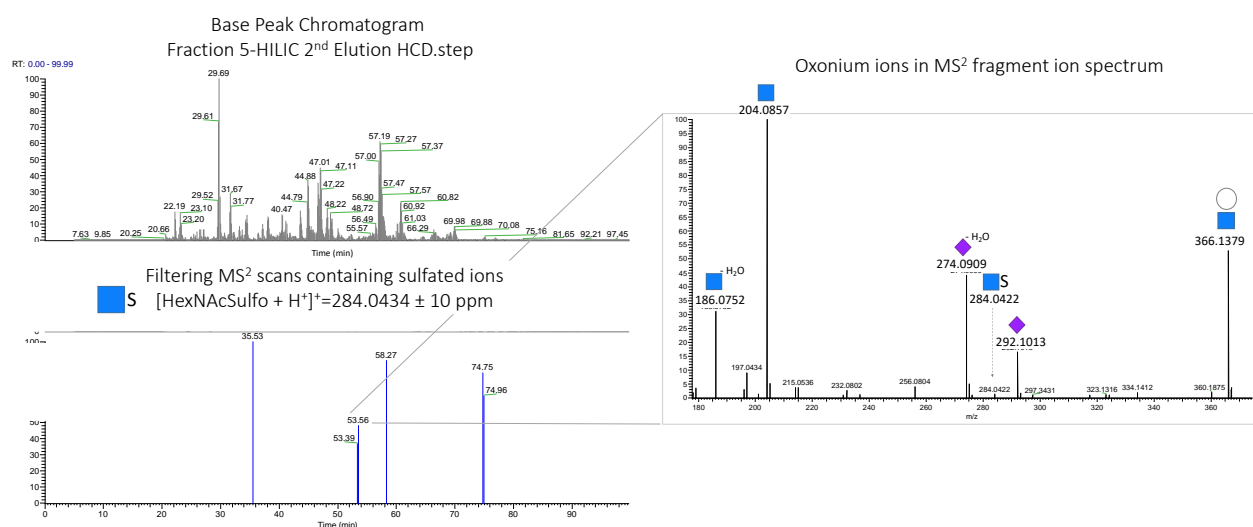
Appendix

A.1. In-depth analysis of the low-abundant *N*-glycoproteome of human blood plasma

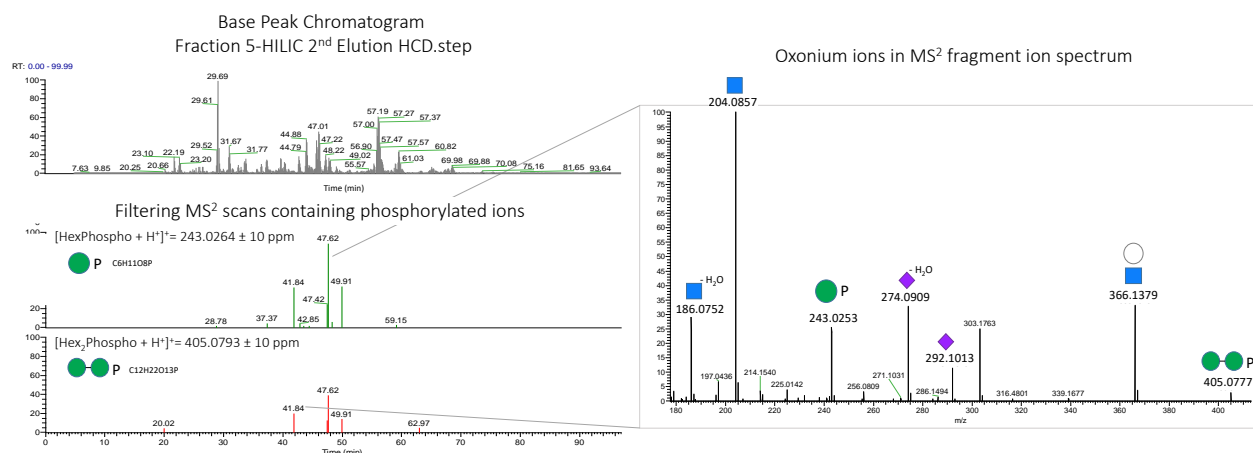
A.1. Tables

All supplementary tables can be found in the file “Supplementary Tables A1.xlsx” that was uploaded to the MPI server and to the cloud (<https://nextcloud.mpi-magdeburg.mpg.de/index.php/s/X8rsEPQRAAHY7Zi>). This supplementary data is also presented in Zuniga-Banuelos et al. (2025a) [228].

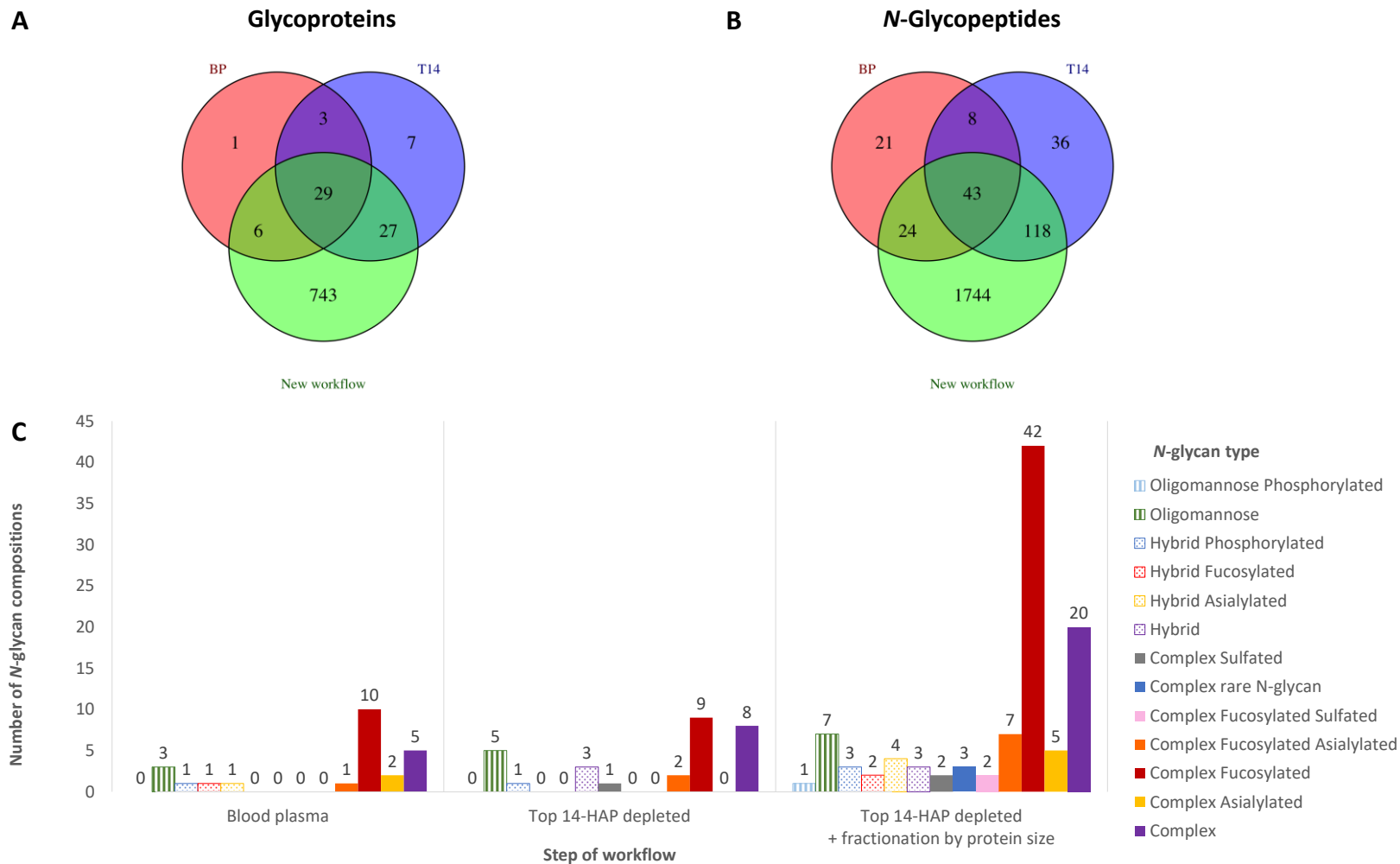
A.1. Figures



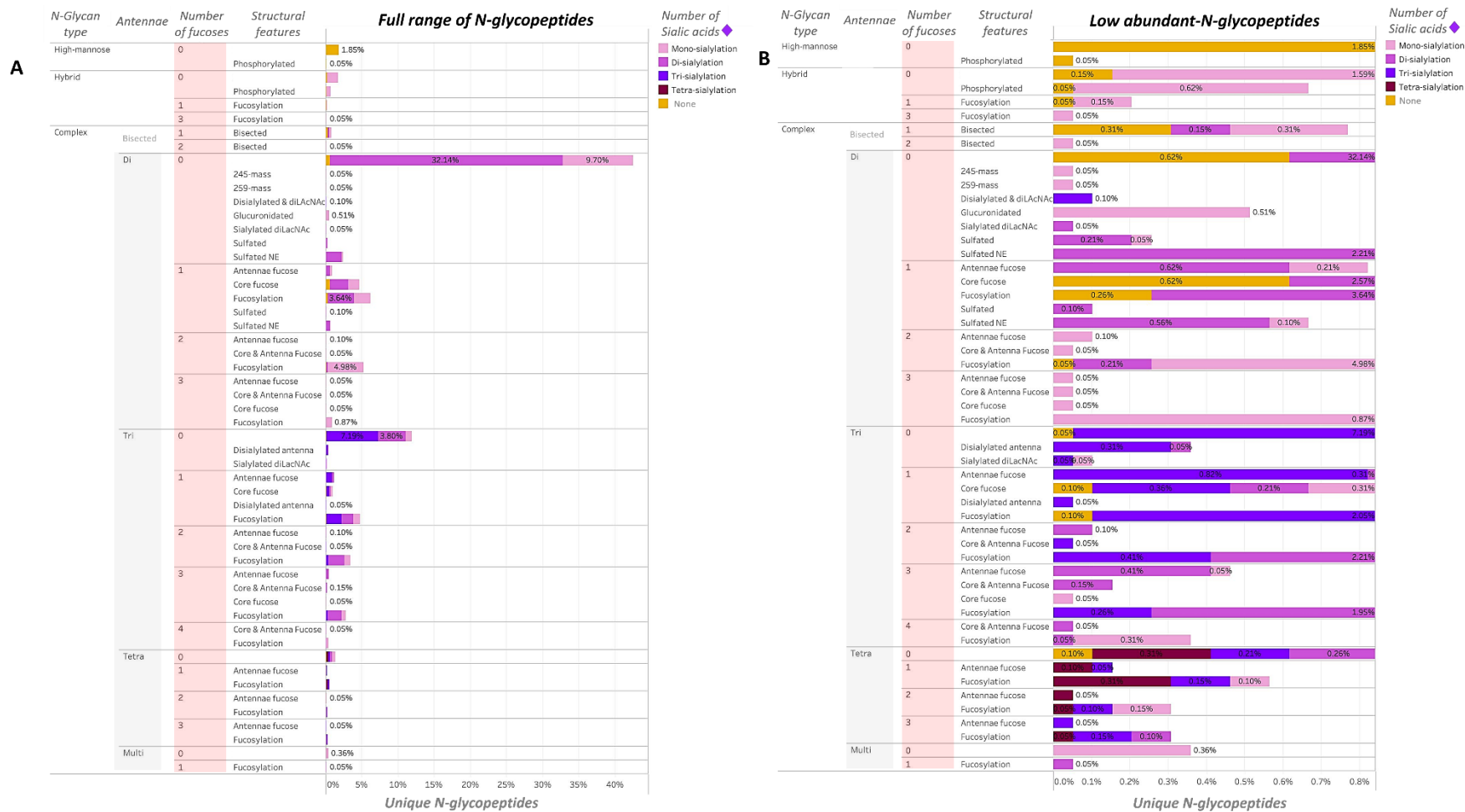
Supplementary Figure A.1.1. Oxonium ion filter applied to identify oxonium ions present in sulfated glycopeptides. Filter configured in Thermo Xcalibur Qual Browser software.



Supplementary Figure A.1.2. Oxonium ion filters applied to identify oxonium ions present in phosphorylated glycopeptides. Filter configured in Thermo Xcalibur Qual Browser software.

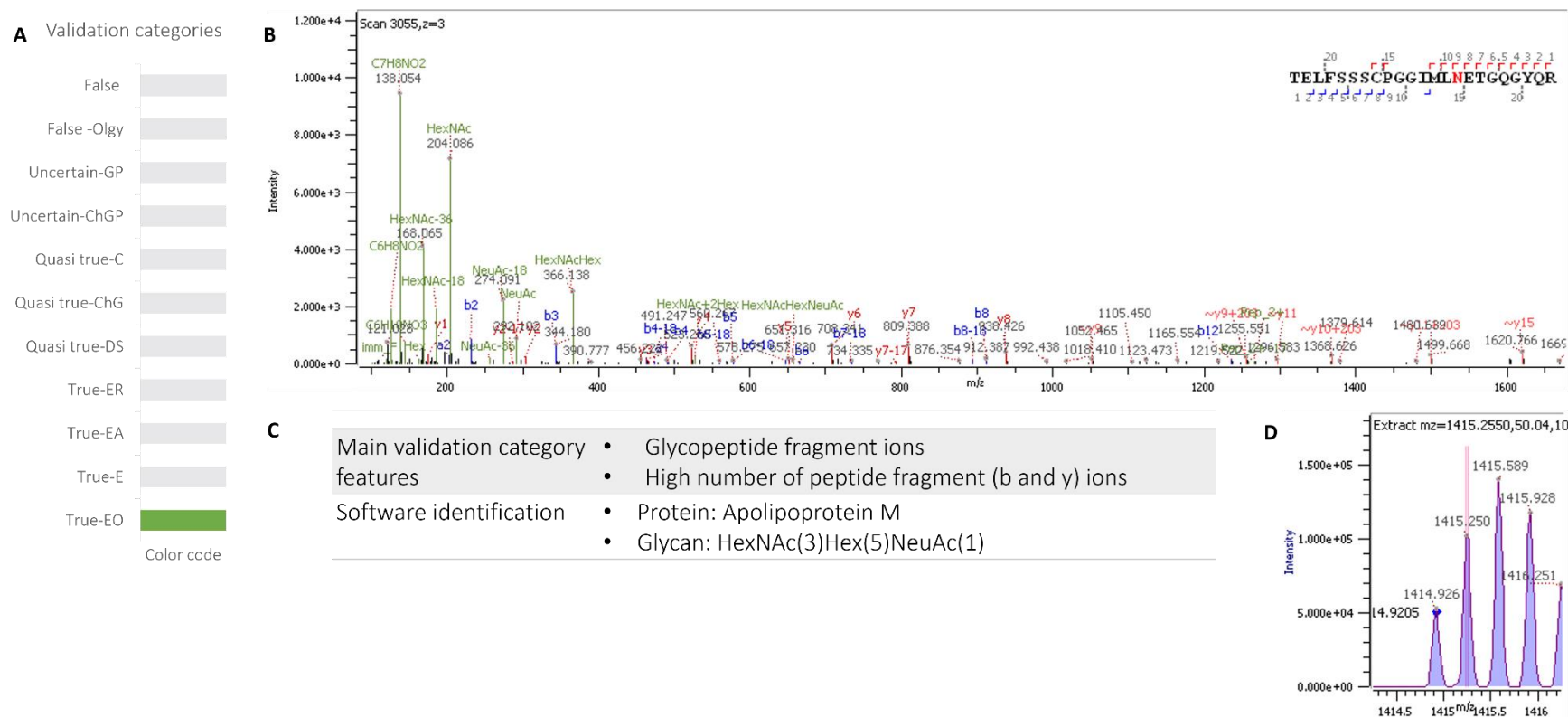


Supplementary Figure A.1.3. Comparison of the gain of the new workflow with respect to the identification of glycoprotein, N-glycopeptide and N-glycan compositions. (A) Number of glycoproteins identified for each step of the sample preparation workflow. (B) Number of N-glycopeptides identified for each step of the sample preparation workflow. (C) The distribution of N-glycan compositions per N-glycan type is shown for each workflow step. BP: Blood plasma. T14: Top 14-HAP depleted. New workflow: the here established sample preparation workflow. The N-glycan types known as “Hybrid” and “Complex” refer to compositions including sialic acid. The N-glycan types, which compositions does not include sialic acid, are specified as “Asialylated”.



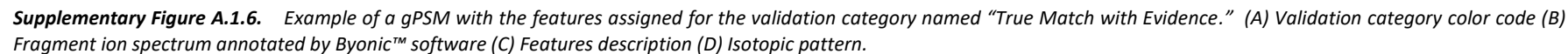
Supplementary Figure A.1.4. N-glycopeptides identified in blood plasma after in-depth N-glycoproteomic analysis. (A) Distribution of unique N-glycopeptides according to N-glycan attached. The x-axis is in full range. (B) Closer overview of the distribution of glycopeptides bearing modified or rare N-glycans.

True Match with Outstanding Evidence

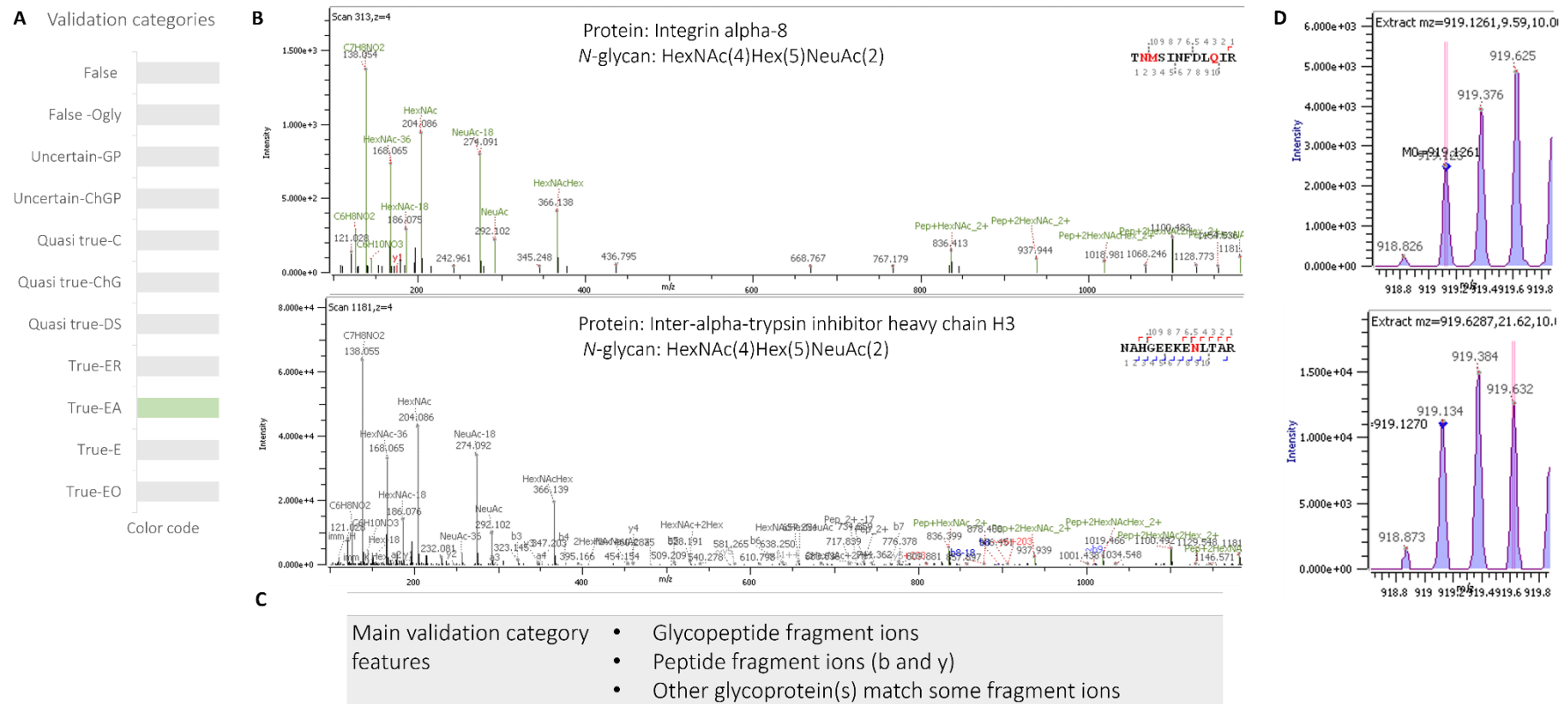


Supplementary Figure A.1.5. Example of a gPSM with the features assigned for the validation category named “True Match with Outstanding Evidence.” (A) Validation category color code (B) Fragment ion spectrum annotated by Byonic™ software (C) Features description (D) Isotopic pattern.

141

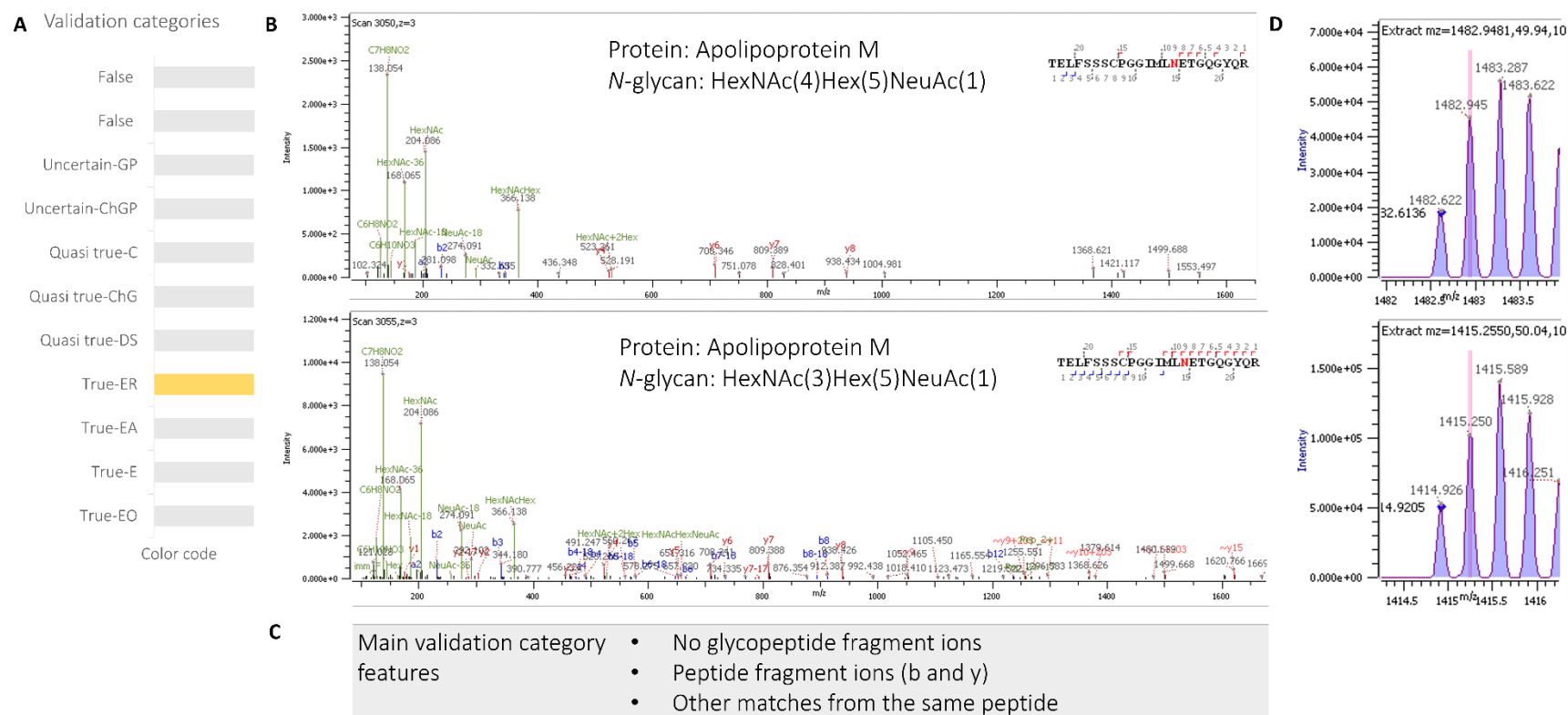


True Match with Evidence and Alternatives



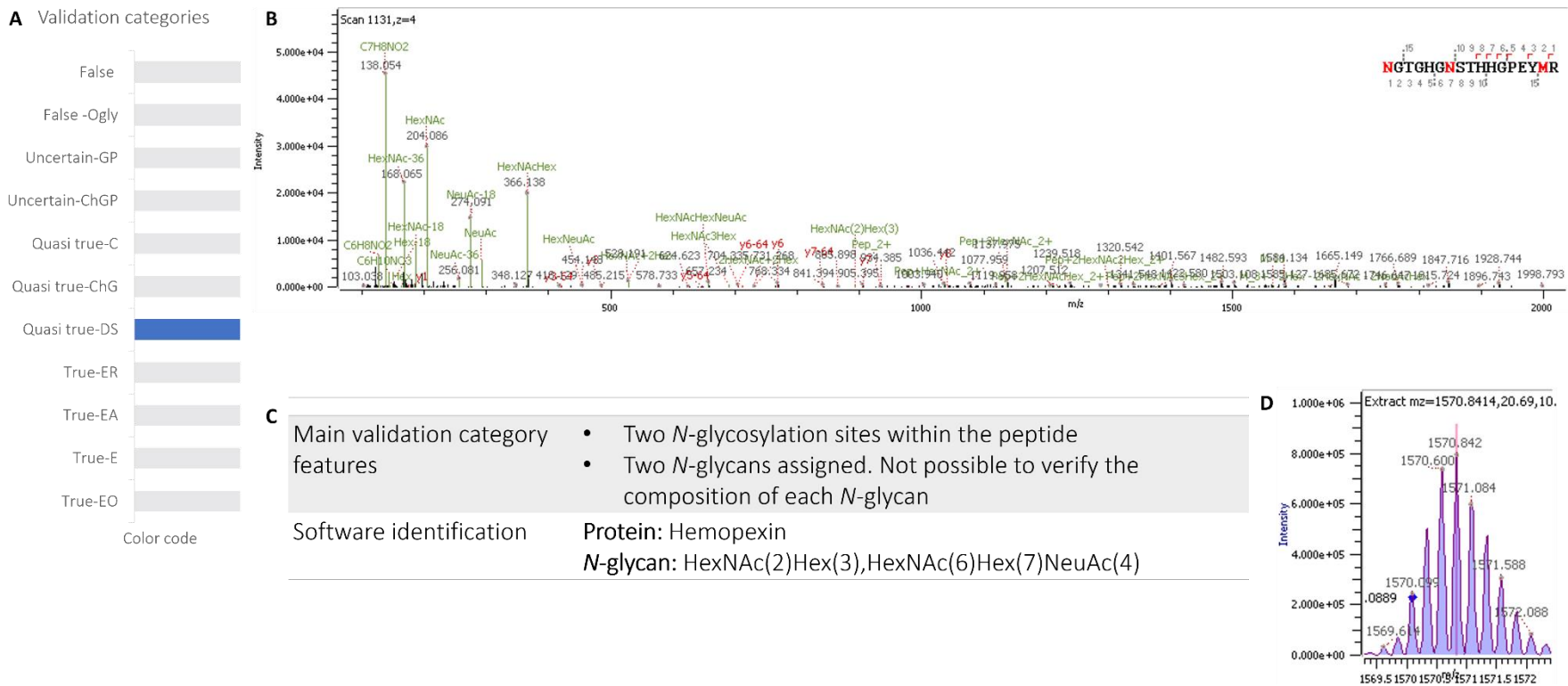
Supplementary Figure A.1.7. Example of a gPSM with the features assigned for the validation category named “True Match with Evidence and Alternatives.” (A) Validation category color code (B) Fragment ion spectrum annotated by Byonic™ software (C) Features description (D) Isotopic pattern.

True Match with Evidence Related to matches from the same peptide



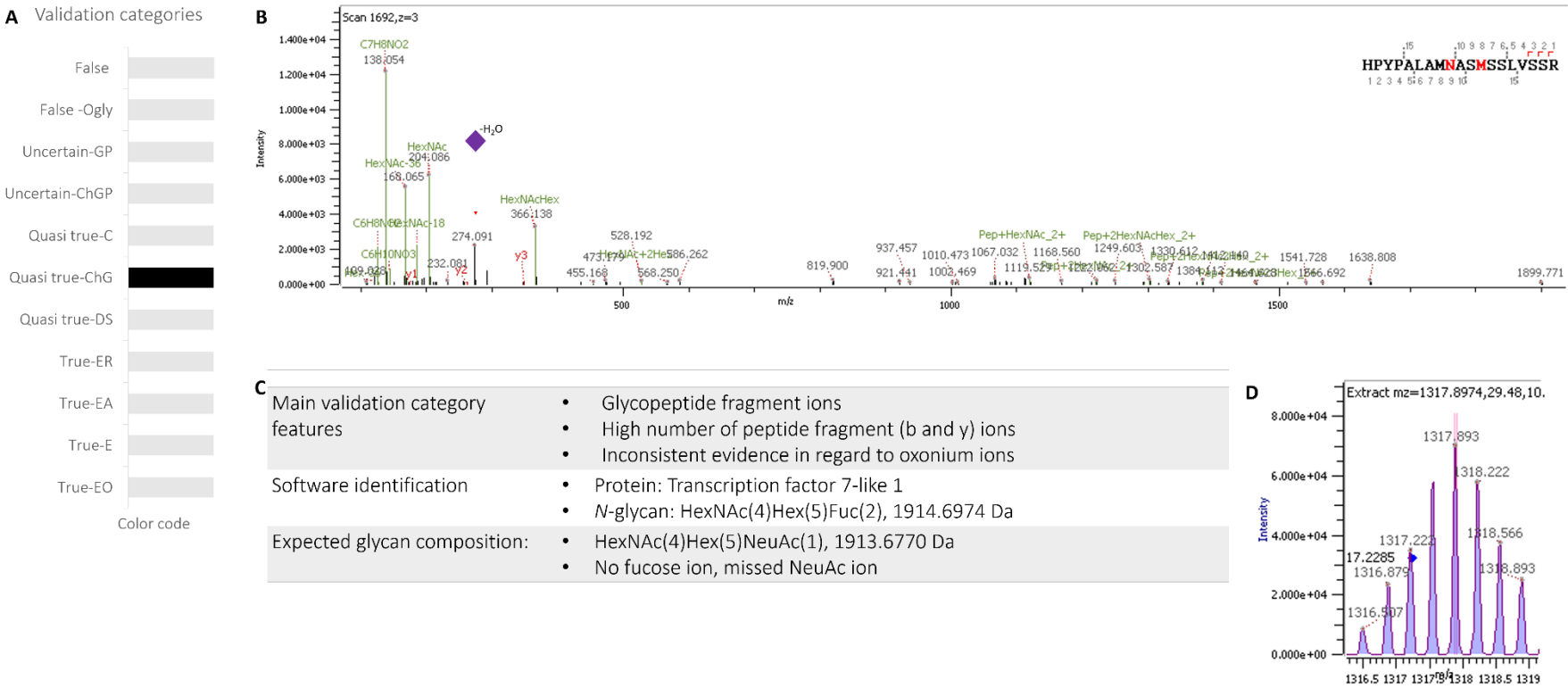
Supplementary Figure A.1.8. Example of a gPSM with the features assigned for the validation category named as “True Match with Evidence Related to matches from the same peptide.”
(A) Validation category color code (B) Fragment ion spectra annotated by Byonic™ software (C) Features description (D) Isotopic pattern.

Quasi-true Match Double Site



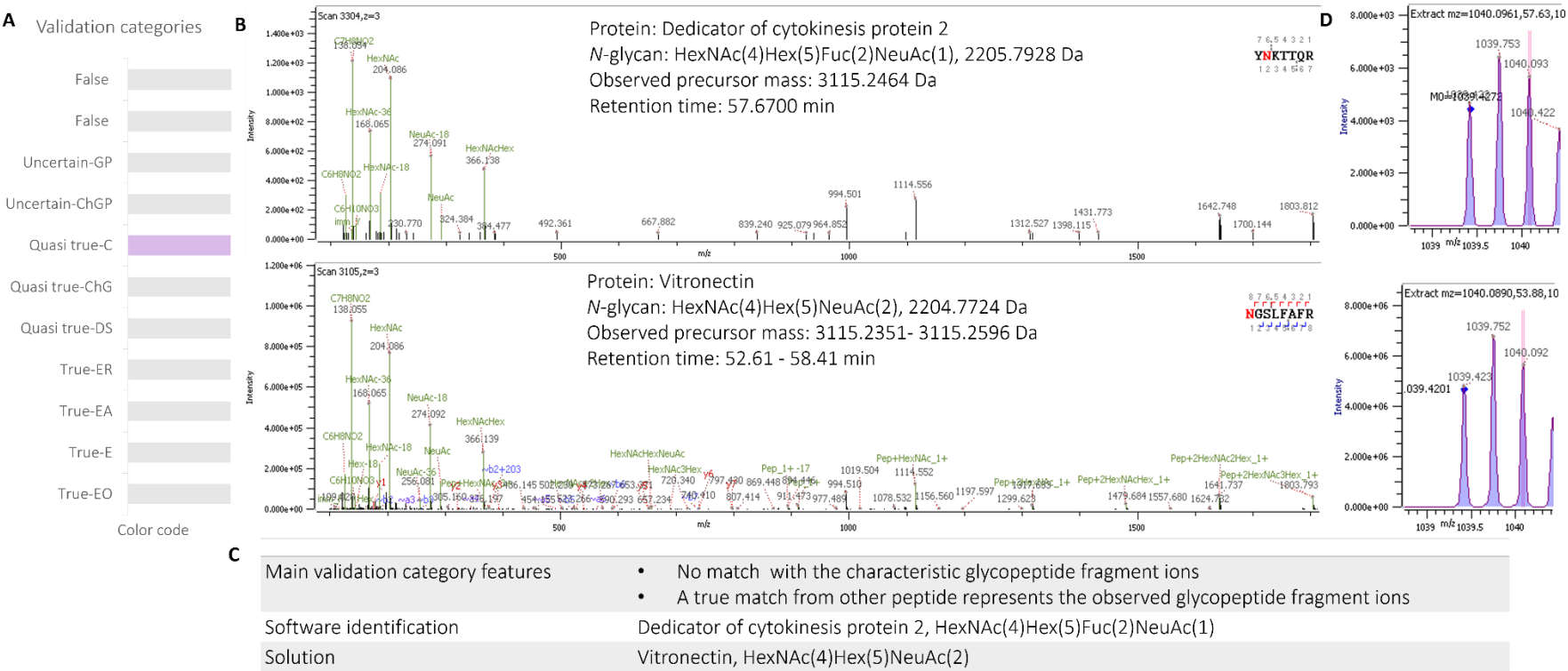
Supplementary Figure A.1.9. Example of a gPSM with the features assigned for the validation category named “Quasi-true Match Double Site.” (A) Validation category color code (B) Fragment ion spectrum annotated by Byonic™ software (C) Features description (D) Isotopic pattern.

Quasi-true Match change glycan



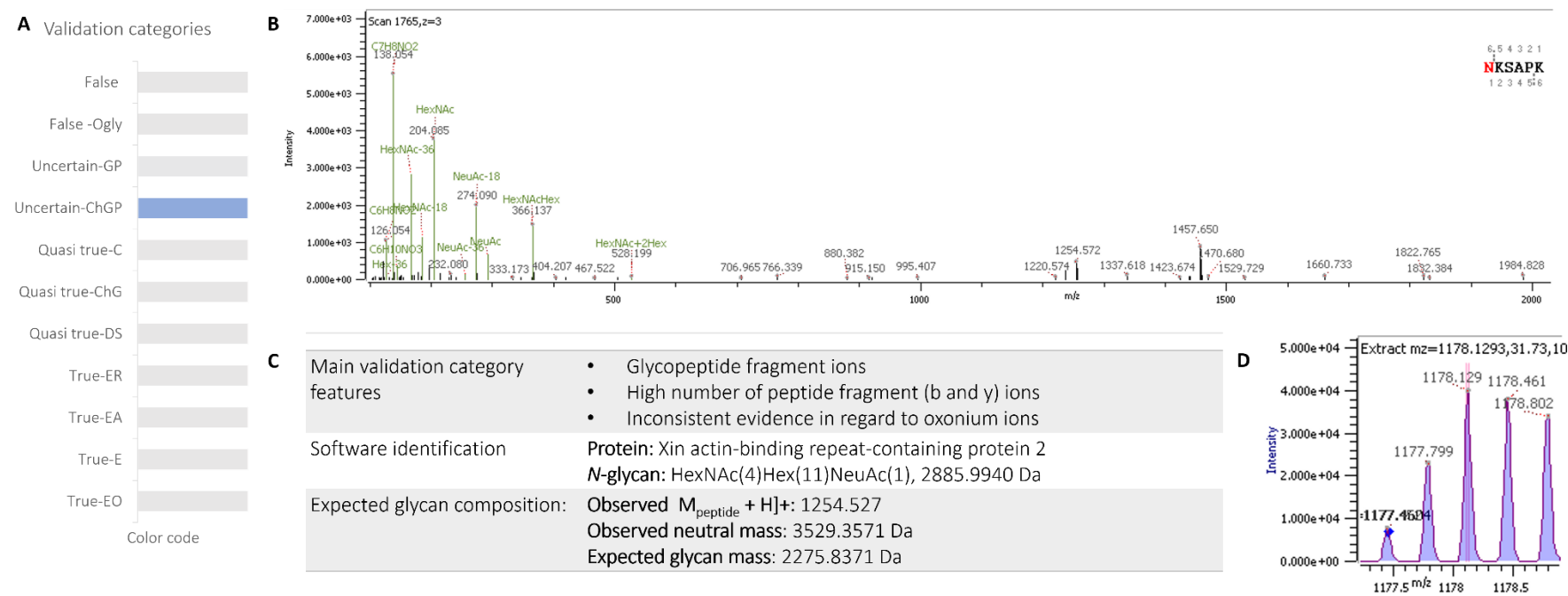
Supplementary Figure A.1.10. Example of a gPSM with the features assigned for the validation category named “Quasi-true Match change glycan.” (A) Validation category color code (B) Fragment ion spectrum annotated by Byonic™ software. (C) Features description (D) Isotopic pattern.

Quasi-true Corrected Match



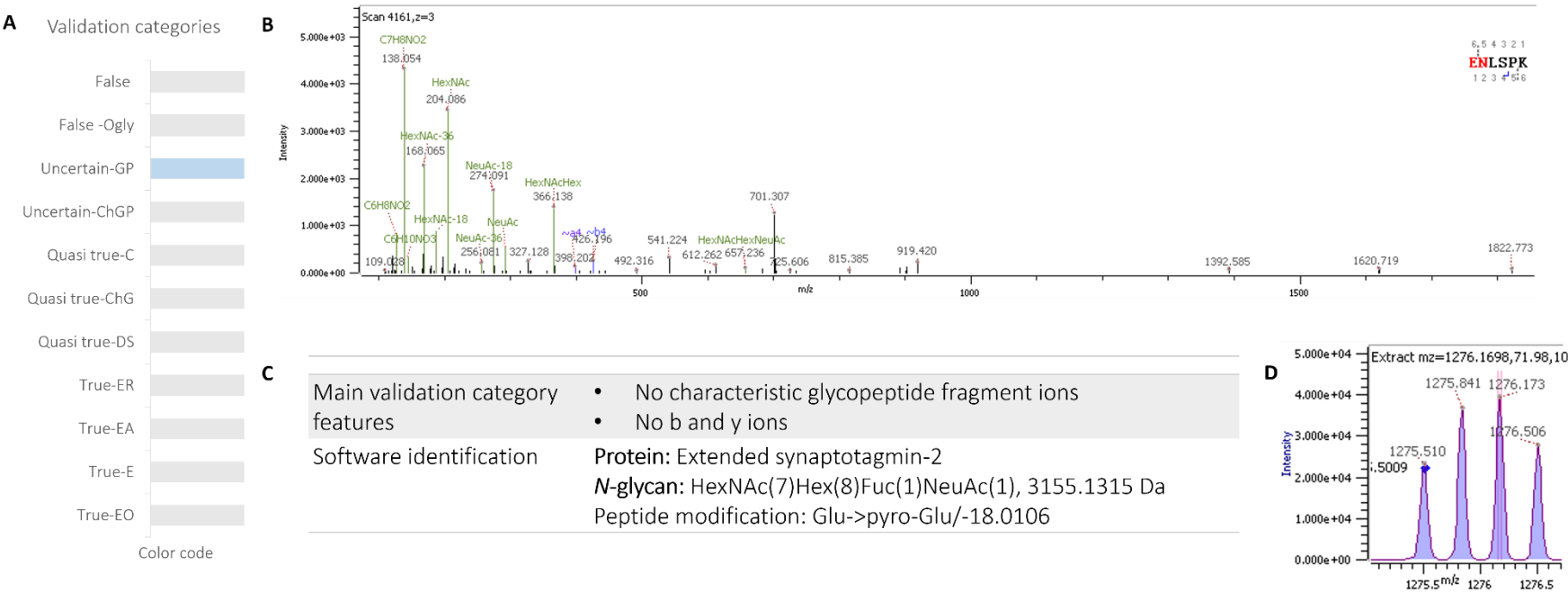
Supplementary Figure A.1.11. Example of a gPSM with the features assigned for the validation category named “Quasi-true corrected Match.” (A) Validation category color code (B) Fragment ion spectra annotated by Byonic™ software (C) Features description (D) Isotopic pattern.

Uncertain-Change glycopeptide match



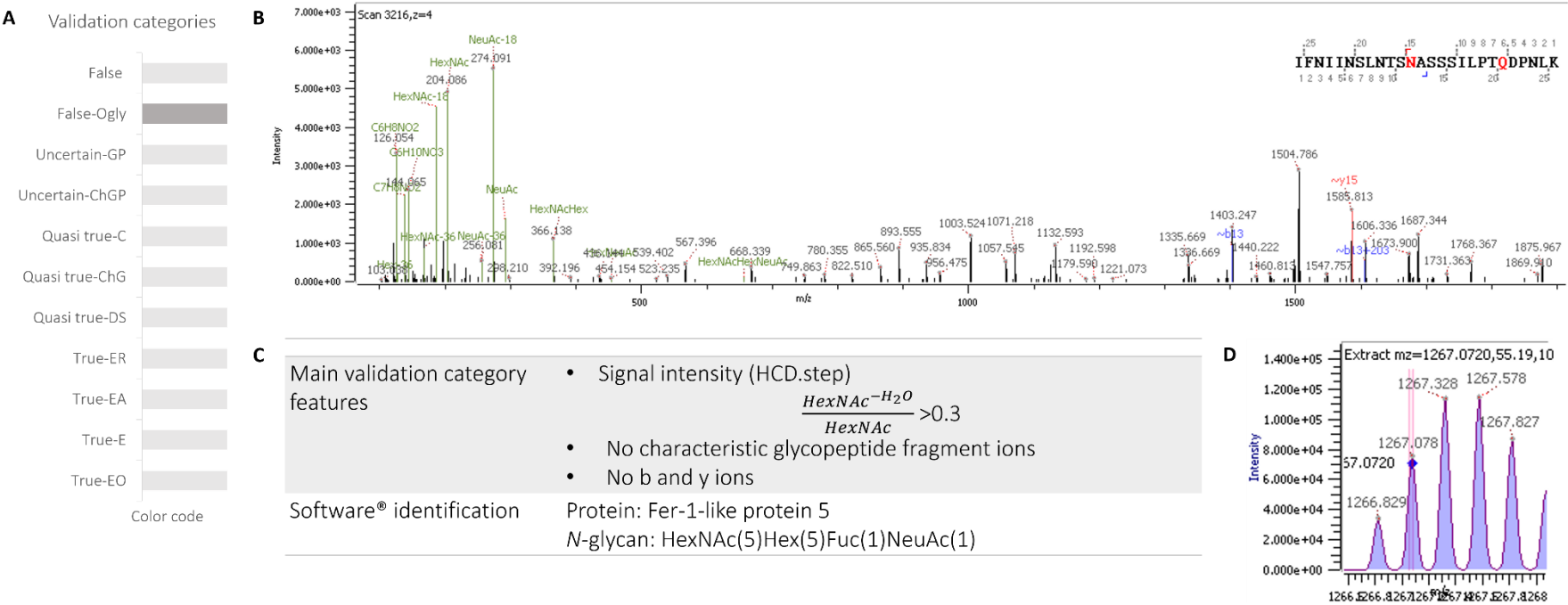
Supplementary Figure A.1.12. Example of a gPSM with the features assigned for the validation category named “Uncertain-Change glycopeptide match.” (A) Validation category color code (B) Fragment ion spectrum annotated by Byonic™ software (C) Features description (D) Isotopic pattern.

Uncertain-glycopeptide



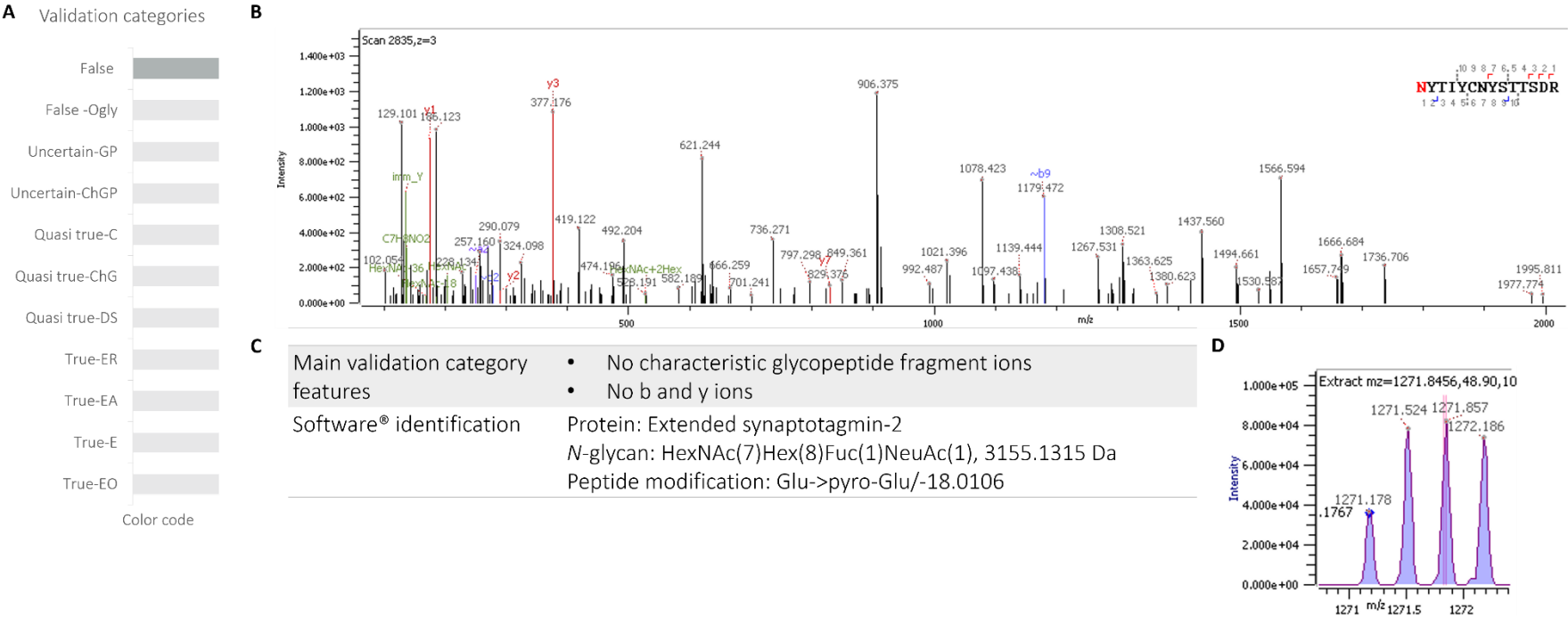
Supplementary Figure A.1.13. Example of a gPSM with the features assigned for the validation category named “Uncertain glycopeptide.” (A) Validation category color code (B) Fragment ion spectrum annotated by Byonic™ software (C) Features description (D) Isotopic pattern.

False O-glycan



Supplementary Figure A.1.14. Example of a gPSM with the features assigned for the validation category named “False O-glycan.” (A) Validation category color code (B) Fragment ion spectrum annotated by Byonic™ software (C) Features description (D) Isotopic pattern.

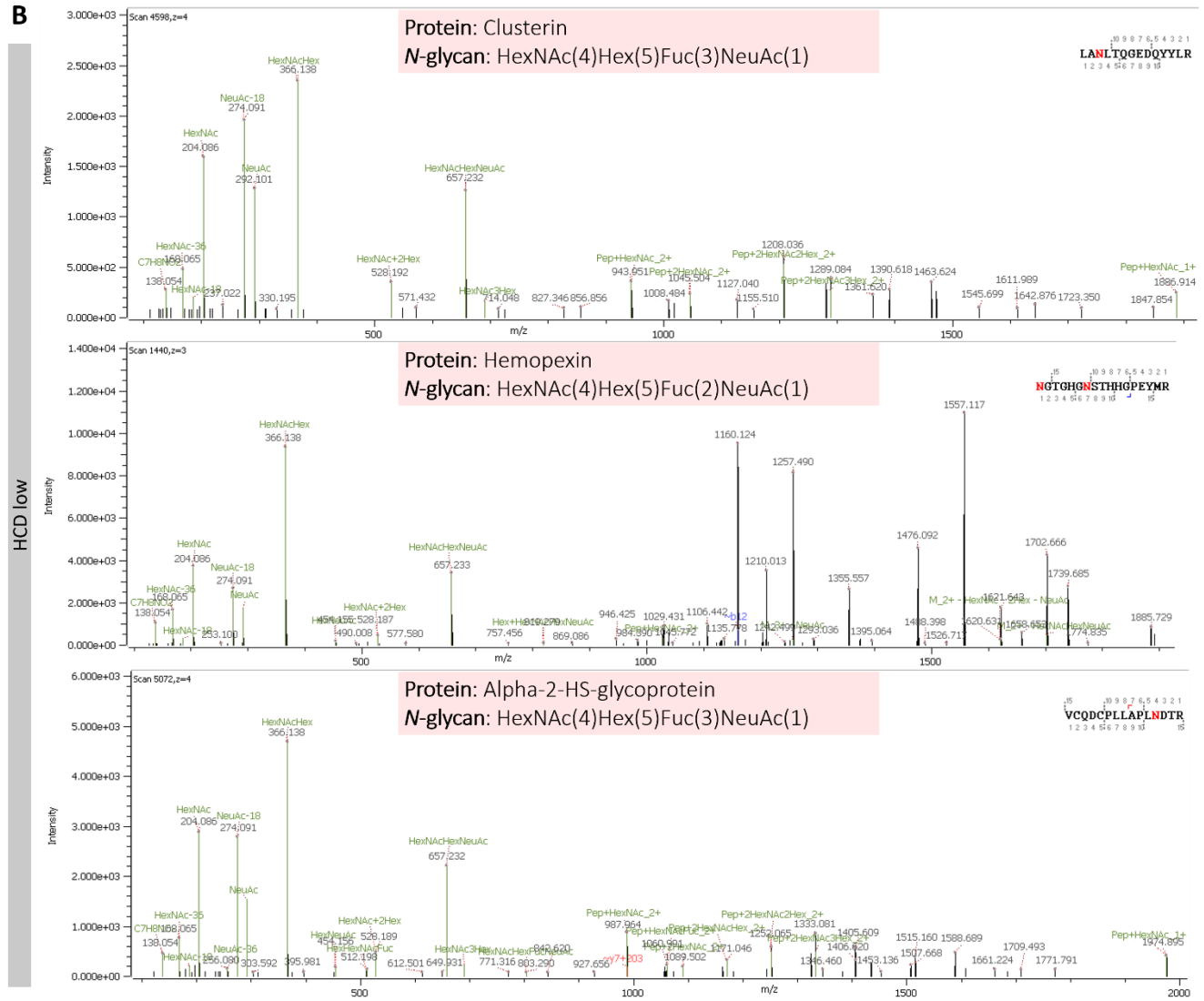
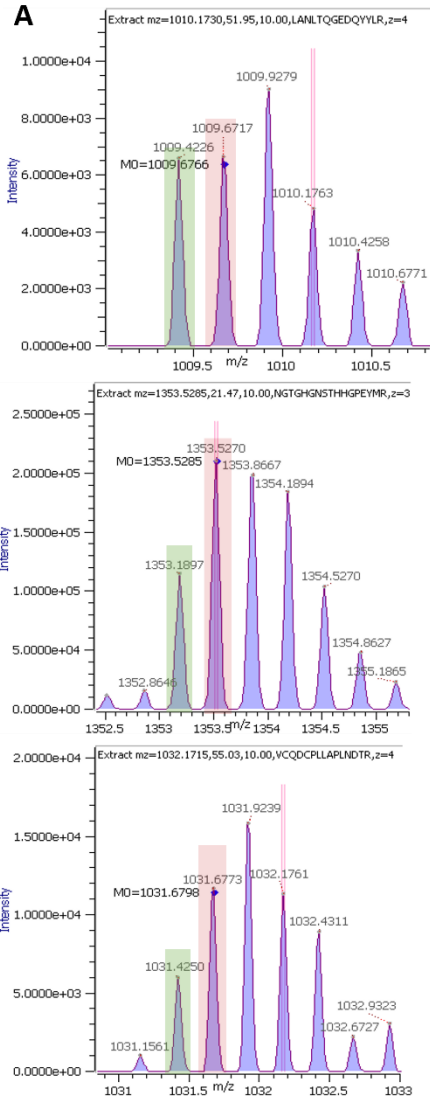
False



Supplementary Figure A.1.15. Example of a gPSM with the features assigned for the validation category named “False.” (A) Validation category color code (B) Fragment ion spectrum annotated by Byonic™ software (C) Features description (D) Isotopic pattern.

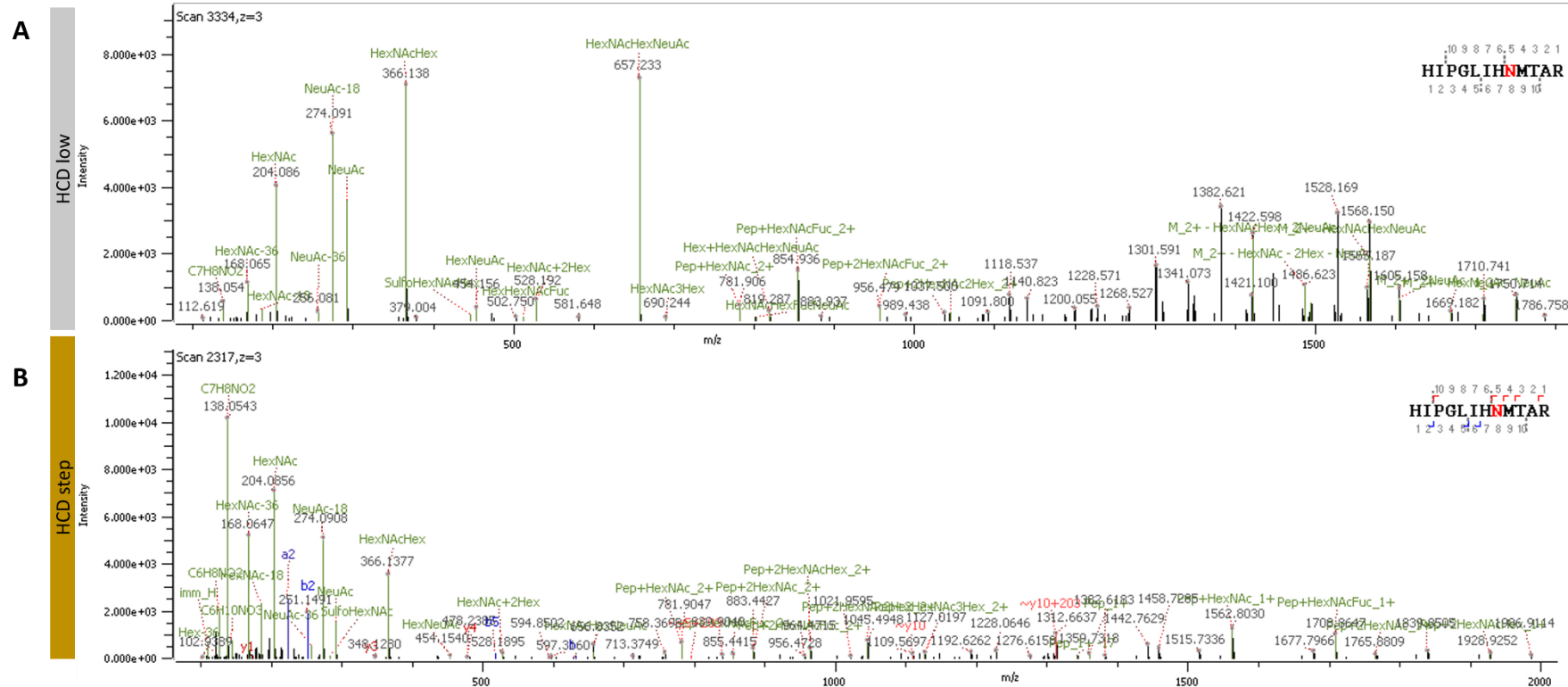
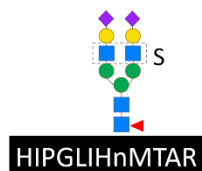


151



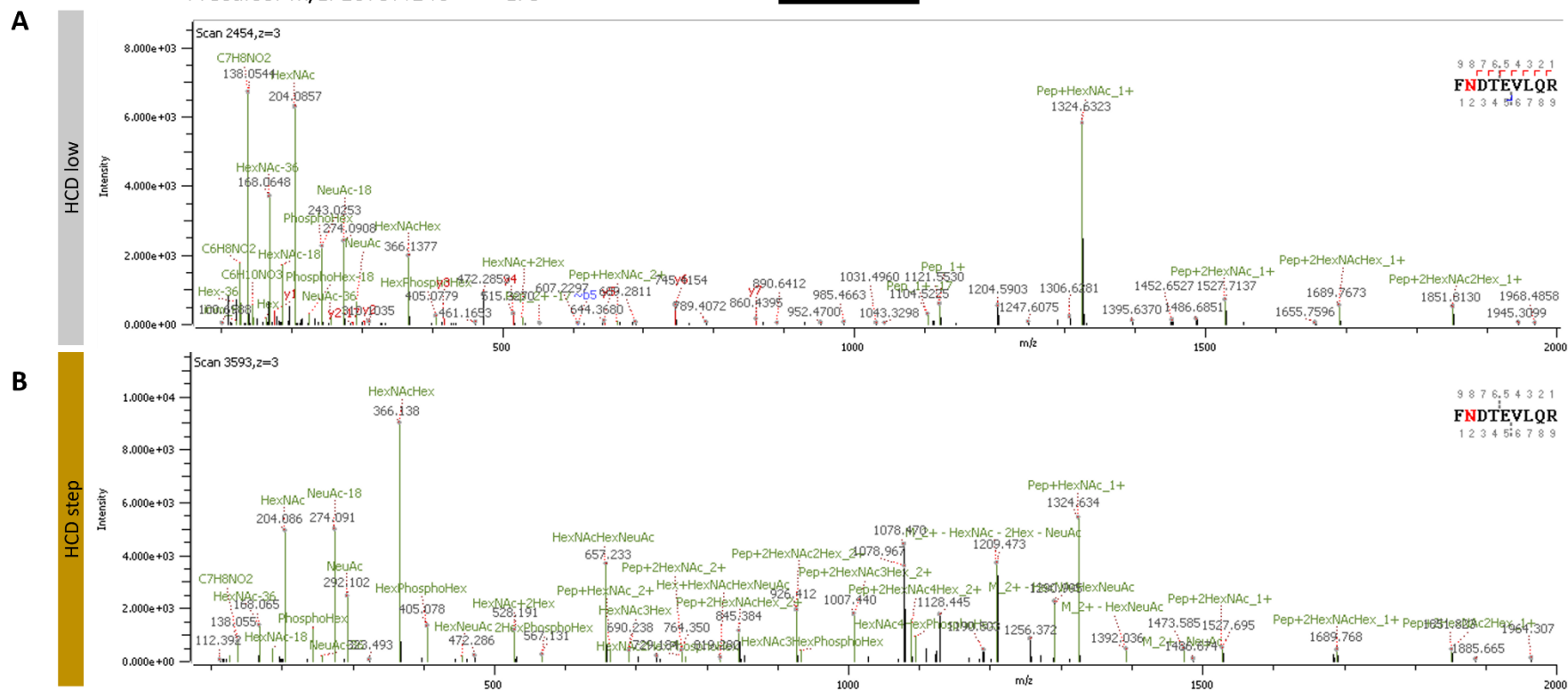
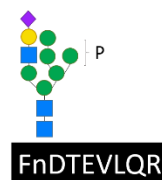
Supplementary Figure A.1.17. Additional examples of gPSM including multiple fucose in a complex N-glycans with a potential error assigning the monoisotopic peak. (A) MS^1 Isotopic distributions. Monoisotopic peaks selected by Byonic™ software (red). (B) MS^2 HCD.low annotation proposed by the software assigning an incorrect N-glycan. Correct precursor ion highlighted in green.

Protein: Extracellular matrix protein 1
 N-glycan: HexNAc(4)Hex(5)Fuc(1)NeuAc(2)Sulfo(1)
 N-glycosylation site: N-444
 Precursor m/z: 1264.1791 z:3

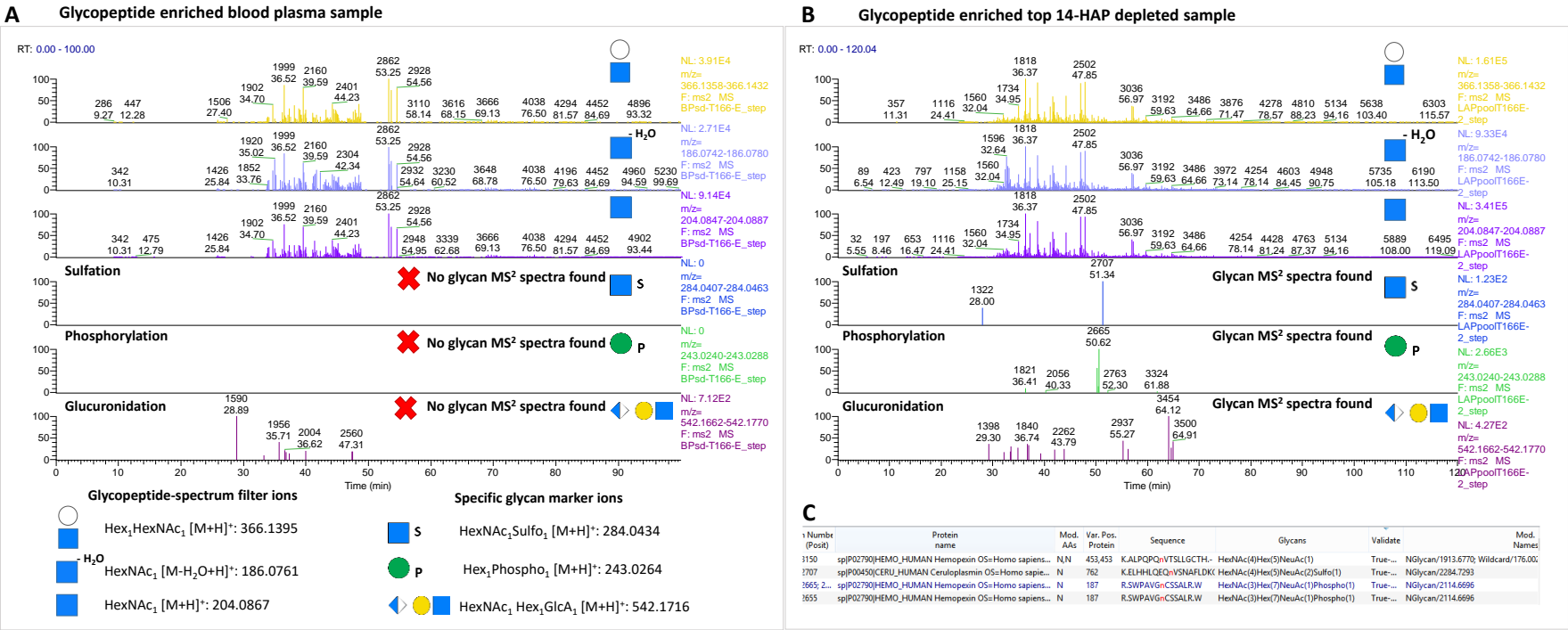


Supplementary Figure A.1.18. Byonic™ software annotation of fragment ions released from sulfated N-glycopeptide. (A) HCD.low fragment ion spectrum (B) HCD.step fragment ion spectrum.

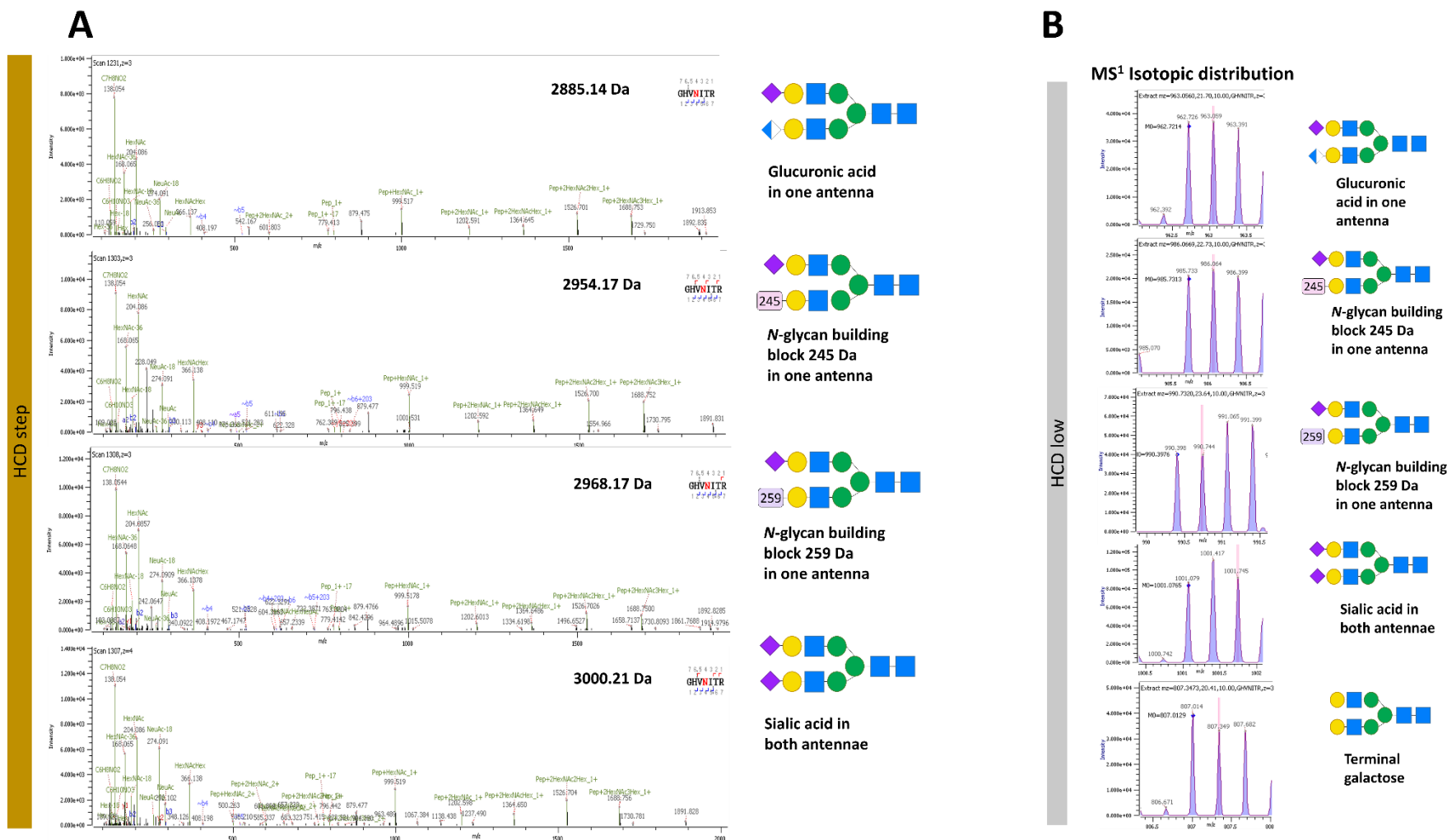
Protein: Biotinidase
 N-glycan: HexNAc(3)Hex(7) NeuAc(1)Phospho(1)
 N-glycosylation site: N-150
 Precursor m/z: 1079.4146 z: 3



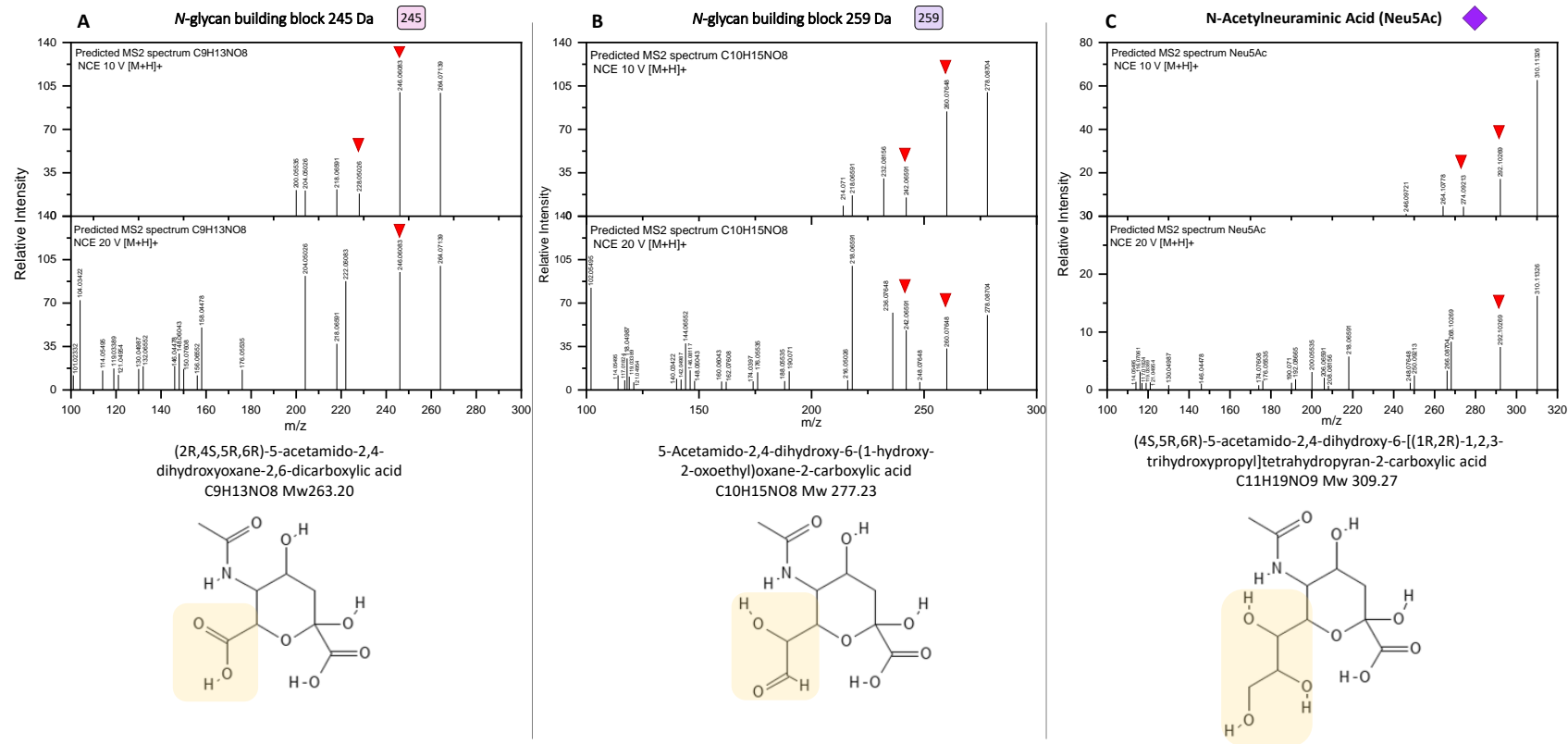
Supplementary Figure A.1.19. Byonic™ software annotation of fragment ions released from phosphorylated N-glycopeptide. (A) HCD.low fragment ion spectrum (B) HCD.step fragment ion spectrum.



Supplementary Figure A.1.20. Searching for rare N-glycopeptides in untreated blood plasma and top 14-HAP depleted samples. (A) LC-MS/MS-HCD.step analysis of glycopeptide enriched blood plasma. No evidence of sulfated, phosphorylated, or glucuronidated glycopeptide-derived MS² spectra. The signals observed in the glucuronidation marker ion filter do not derive from glycopeptides. (B) LC-MS/MS-HCD.step analysis of glycopeptide enriched of top 14-HAP depleted blood plasma. Evidence of sulfated, phosphorylated, or glucuronidated glycopeptide-derived MS² spectra. (C) Top 14-HAP depleted blood plasma Byonic™ search identified the N-glycopeptides holding sulfated, phosphorylated, and glucuronidated N-glycans. The layout filters containing glycan marker ions were created in Thermo Xcalibur Qual Browser (Version 2.2, Thermo Scientific) software.



Supplementary Figure A.1.21. Byonic™ software annotation of HCD.step fragment ion spectra containing rare N-glycans found in N121 from prothrombin (UniProtKB P00734). (A) N-glycopeptide annotation by the software. (B) MS¹ Isotopic distribution of precursor ions observed through HCD.low acquisition.



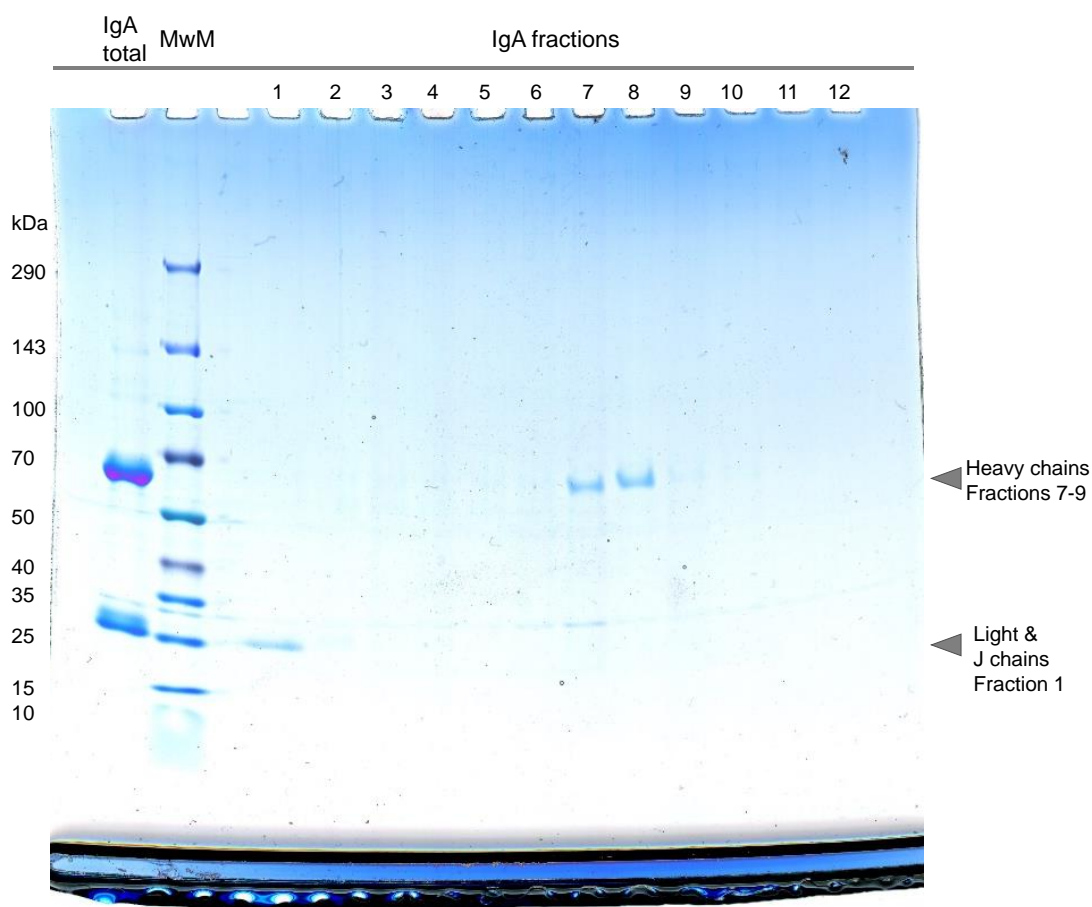
Supplementary Figure A.1.22. ESI-MS/MS prediction for two molecules proposed as the N-glycan building blocks of the rare N-glycopeptide identifications observed. **(A)** Predicted molecule fragmentation with the chemical composition C9H13NO8 suggested for N-glycan building block 245 Da. The molecule found in PubChem (CID 59861591) is shown below the spectra. **(B)** Predicted molecule fragmentation with the chemical composition C10H15NO8 suggested for N-glycan building block 259 Da. The molecule found in PubChem (CID 59173067) is shown below the spectra. **(C)** Predicted fragmentation of Neu5Ac; the structure is shown below the spectra (PubChemCID 439197). Red triangles indicate the fragment ions observed in the fragment ion spectra containing the rare N-glycans and sialic acid. Predicted spectra computed by CFM-ID (<https://cfmid.wishartlab.com/predict>) for two collisional energies (10 eV and 20 eV) in positive ion mode, including [M+H]⁺ as an adduct. Molecules drawn in PubChem Sketcher [256-260].

A.2. Site-specific identification of sulfated and *O*-acetylated *N*-glycans in human serum IgA

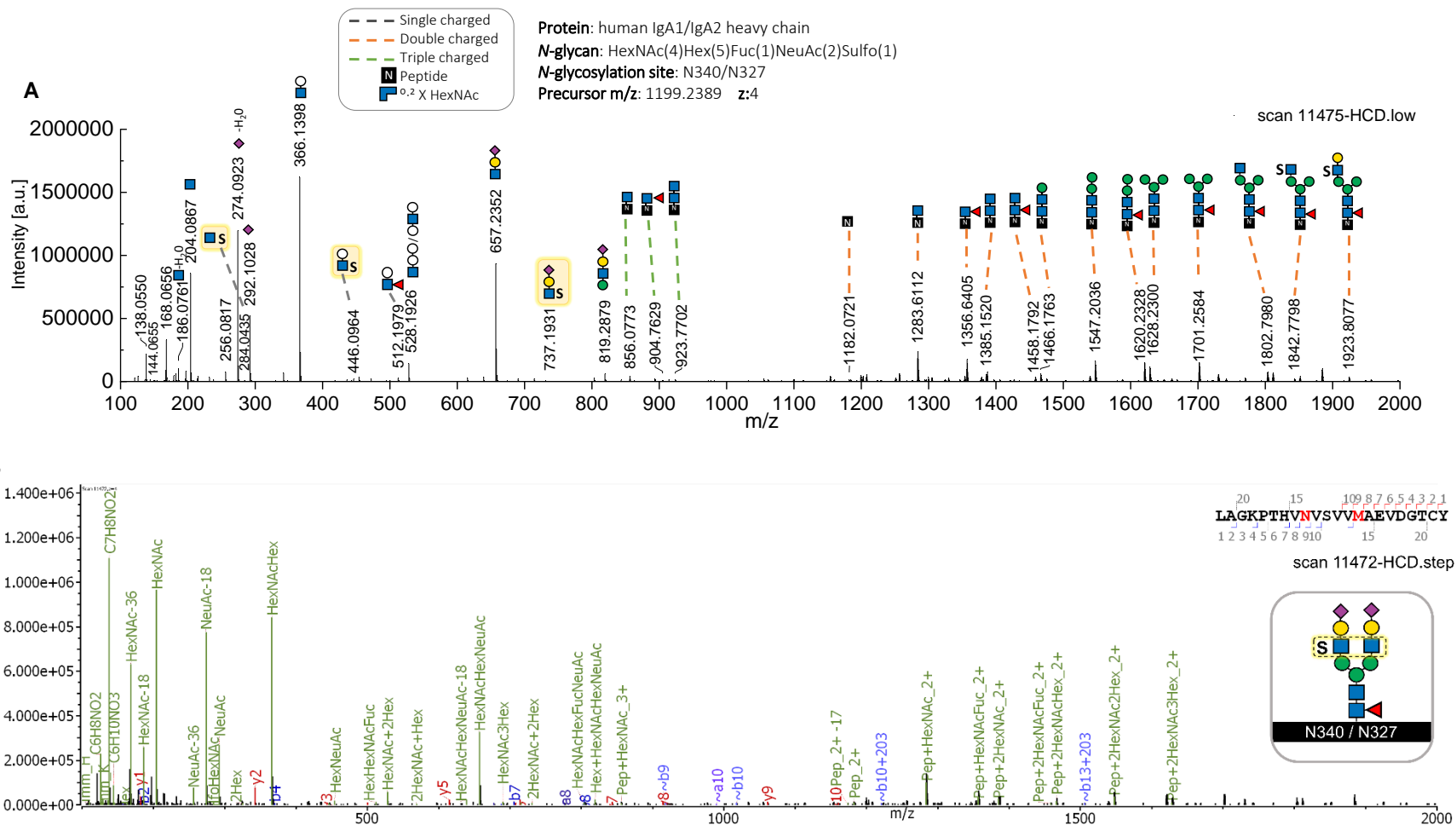
A.2. Tables

All supplementary tables can be found in the file “Supplementary Tables A2.xlsx” that was uploaded to the MPI server and to the cloud (<https://nextcloud.mpi-magdeburg.mpg.de/index.php/s/X8rsEPQRAAHY7Zi>). This supplementary data is also presented in Zuniga-Banuelos et al. (2025b) [229].

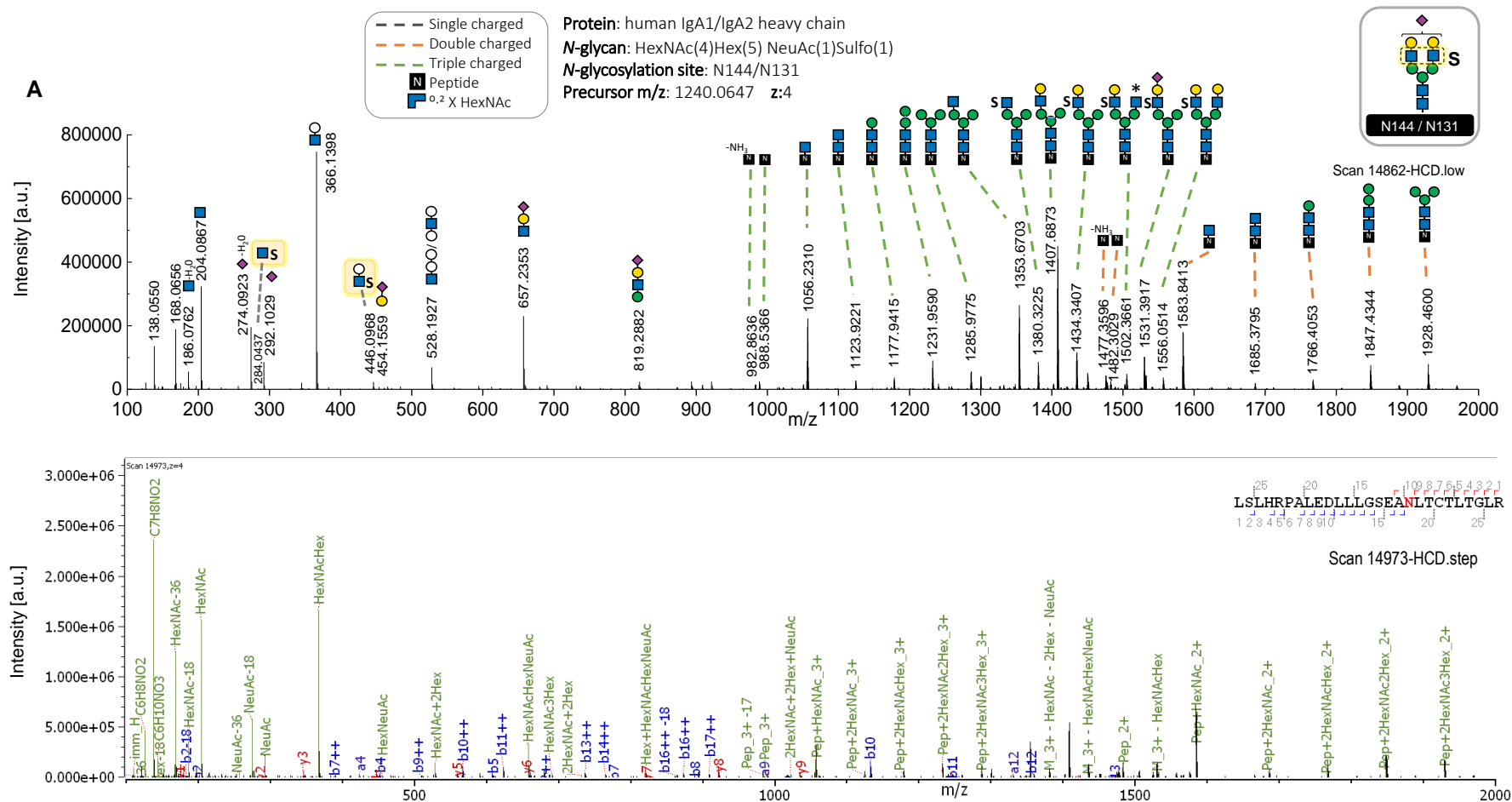
A.2. Figures



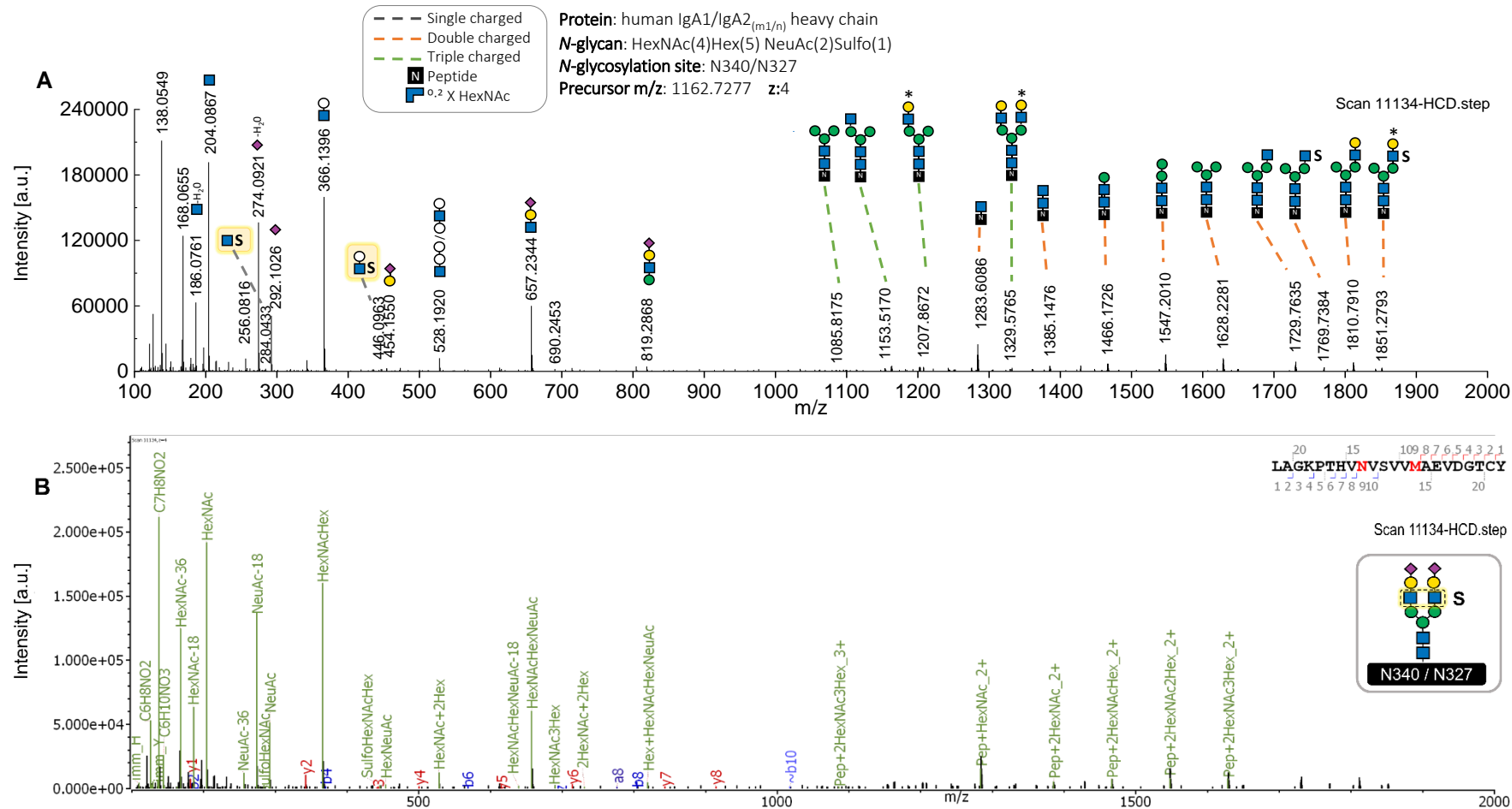
Supplementary Figure A.2.1. Separation of primarily light and heavy chains of human serum IgA subclasses by GELFrEE® system. The protein fractionation was evaluated via SDS-PAGE by loading an aliquot of each IgA fraction. MwM: Molecular weight marker.



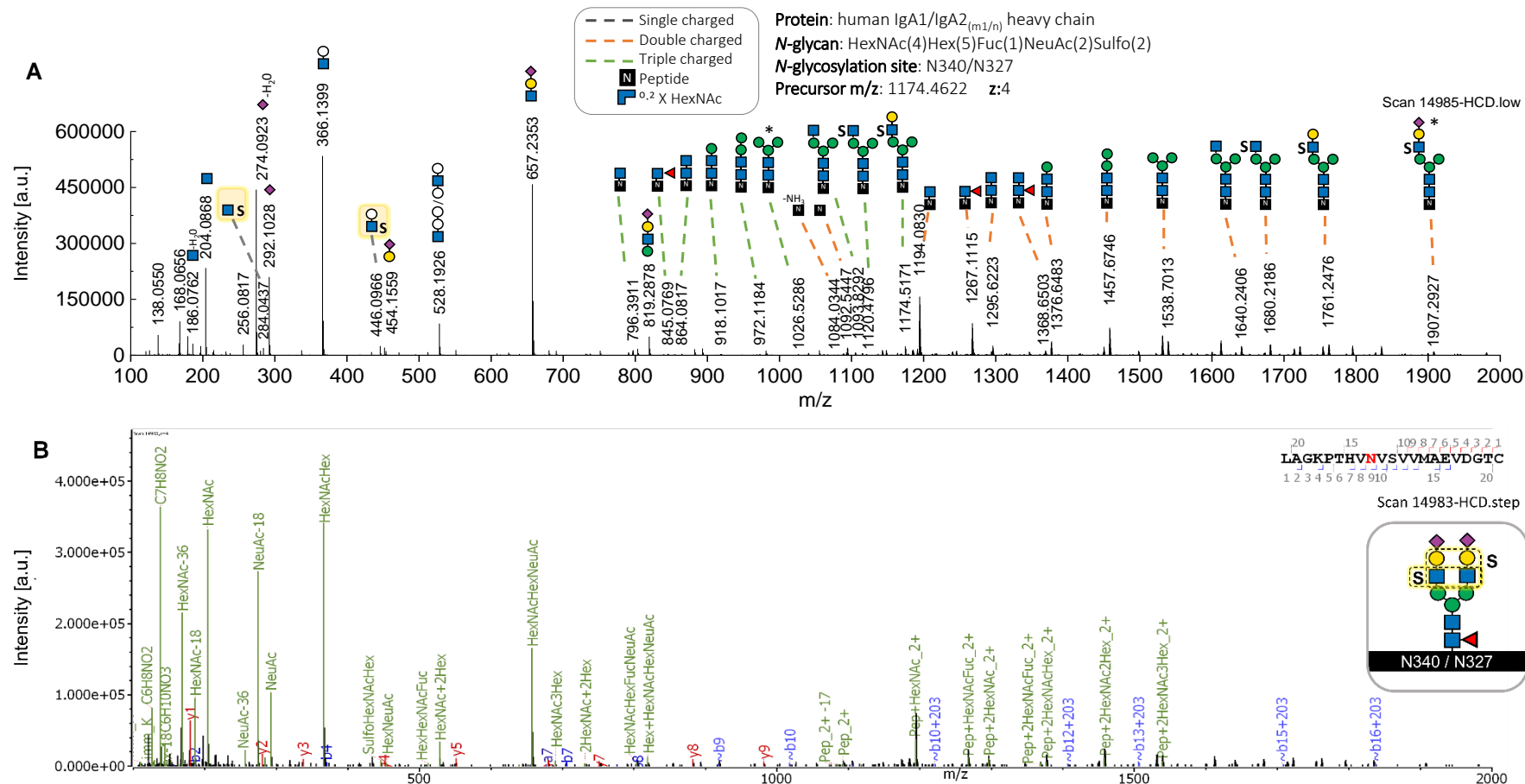
Supplementary Figure A.2.2. Annotated data of a glycopeptide bearing the sulfated N-glycan HexNAc(4)Hex(5)Fuc(1)NeuAc(2)Sulfo(1). (A) Description of the N-glycan moiety through HCD.low spectrum shows B and Y ions. The oxonium marker ion Sulfo1HexNAc1/284.0435 [M+H]⁺ confirms specifically HexNAc sulfation, while the oxonium ion Sulfo1HexNAc1Hex1NeuAc1/737.1931 [M+H]⁺ confirms specifically antenna HexNAc sulfation. (B) The Byonic™ annotation of the HCD.step spectrum refers to a tryptic peptide containing the homologous site N340-IgA1/N327-IgA2(m1/n). The HCD.step spectrum shows the oxonium marker ions confirming HexNAc sulfation. The HCD.low and HCD.step spectra that most extensively describe the glycoform and peptide sequence, acquired within the shortest time, were selected for representing this glycopeptide.



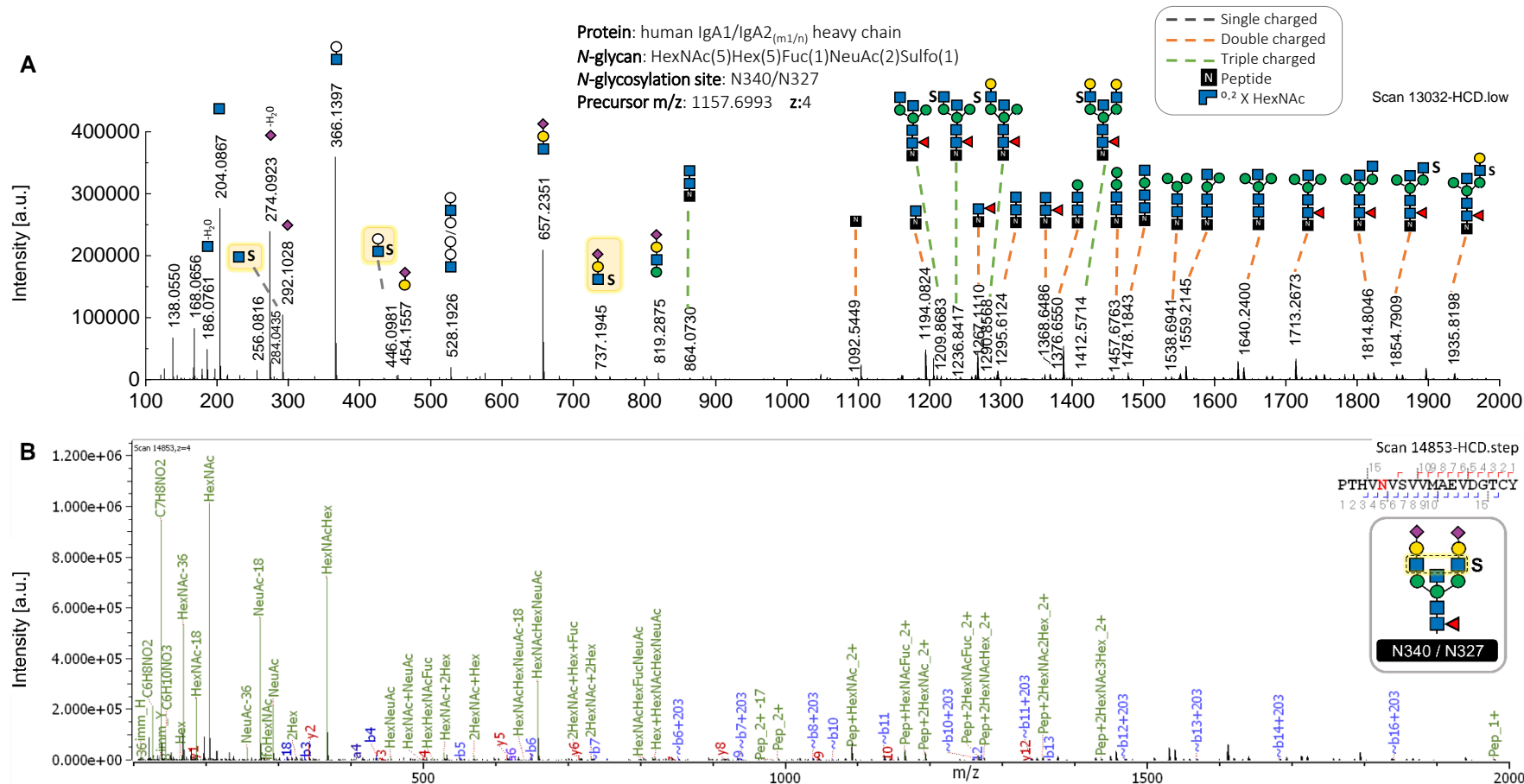
Supplementary Figure A.2.3. Annotated data of a glycopeptide bearing the sulfated N-glycan HexNAc(4)Hex(5)NeuAc(1)Sulfo(1). (A) Description of the N-glycan moiety through HCD.low spectrum shows B and Y ions confirming antenna with a sulfated HexNAc. The N-glycan icon with an asterisk symbol (*) depicts the N-glycan composition by a Y ion assigned to the second isotope peak of the expected m/z. (B) The Byonic™ annotation of the HCD.step spectrum refers to an amino acid sequence of this N-glycopeptide containing the homologous site N144-IgA1/N131-IgA2_(m1/m2/n). The HCD.step spectrum shows the oxonium marker ions confirming HexNAc sulfation (Sulfo₁HexNAc₁/284.0437 [M+H]⁺). The HCD.low and HCD.step spectra that most extensively describe the glycoform and peptide sequence, acquired within the shortest time, were selected for representing this glycopeptide.



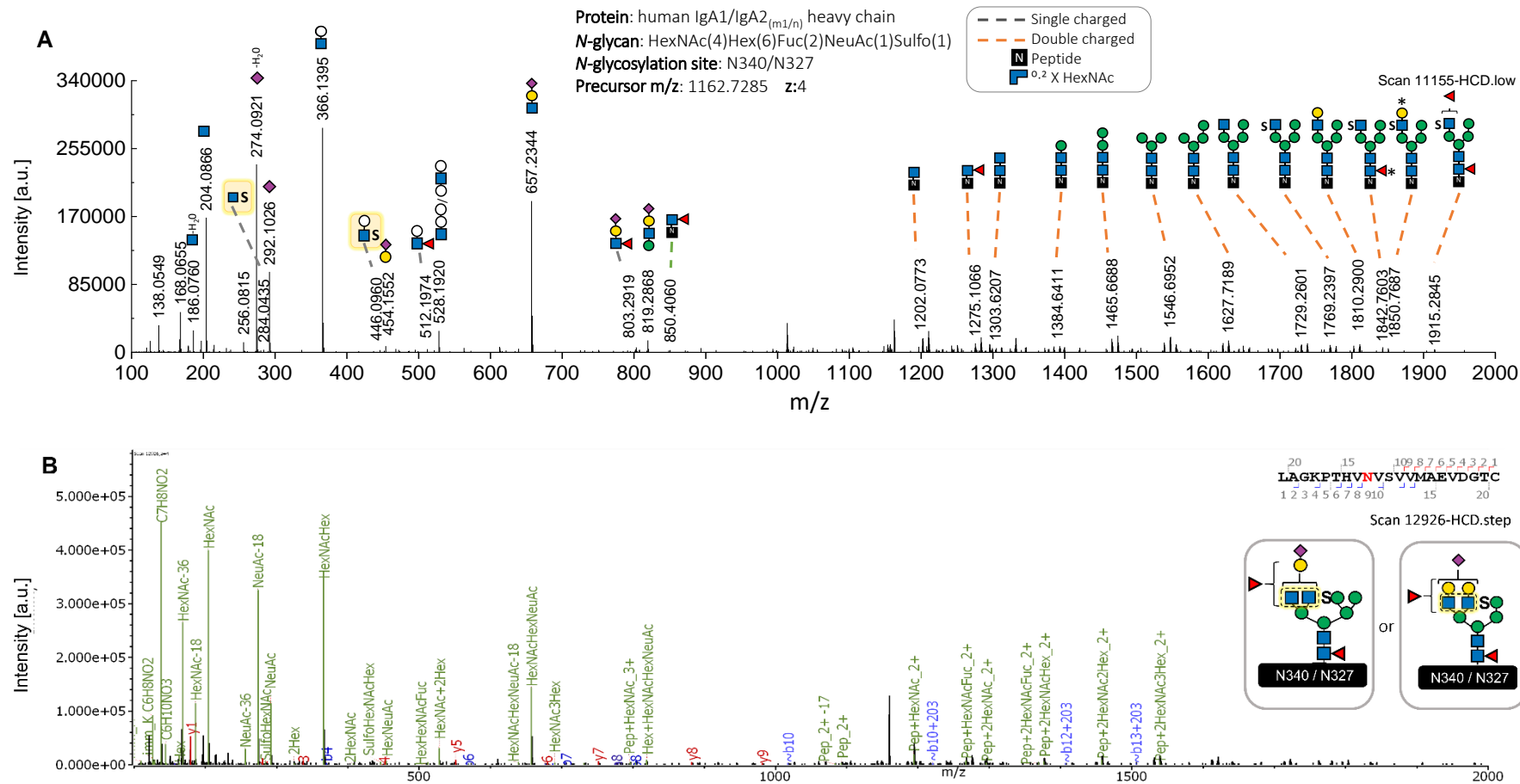
Supplementary Figure A.2.4. Annotated data of a glycopeptide bearing the sulfated N-glycan: HexNAc(4)Hex(5)NeuAc(2)Sulfo(1). (A) Description of the N-glycan moiety through HCD.step spectrum shows B and Y ions confirming antenna with sulfated HexNAc. The N-glycan icon with an asterisk symbol (*) depicts the N-glycan composition by a Y ion assigned to the second isotope peak of the expected m/z. (B) Byonic™ annotation of the HCD.step spectrum refers to an amino acid sequence of this N-glycopeptide containing the homologous site N340-IgA1/N327-IgA2_(m1/n). The HCD.step spectrum shows the oxonium marker ions confirming HexNAc sulfation (Sulfo₁HexNAc₁/284.0433 [M+H]⁺). HCD.step spectrum that most extensively describes the glycoform and peptide sequence was selected for representing this glycopeptide.



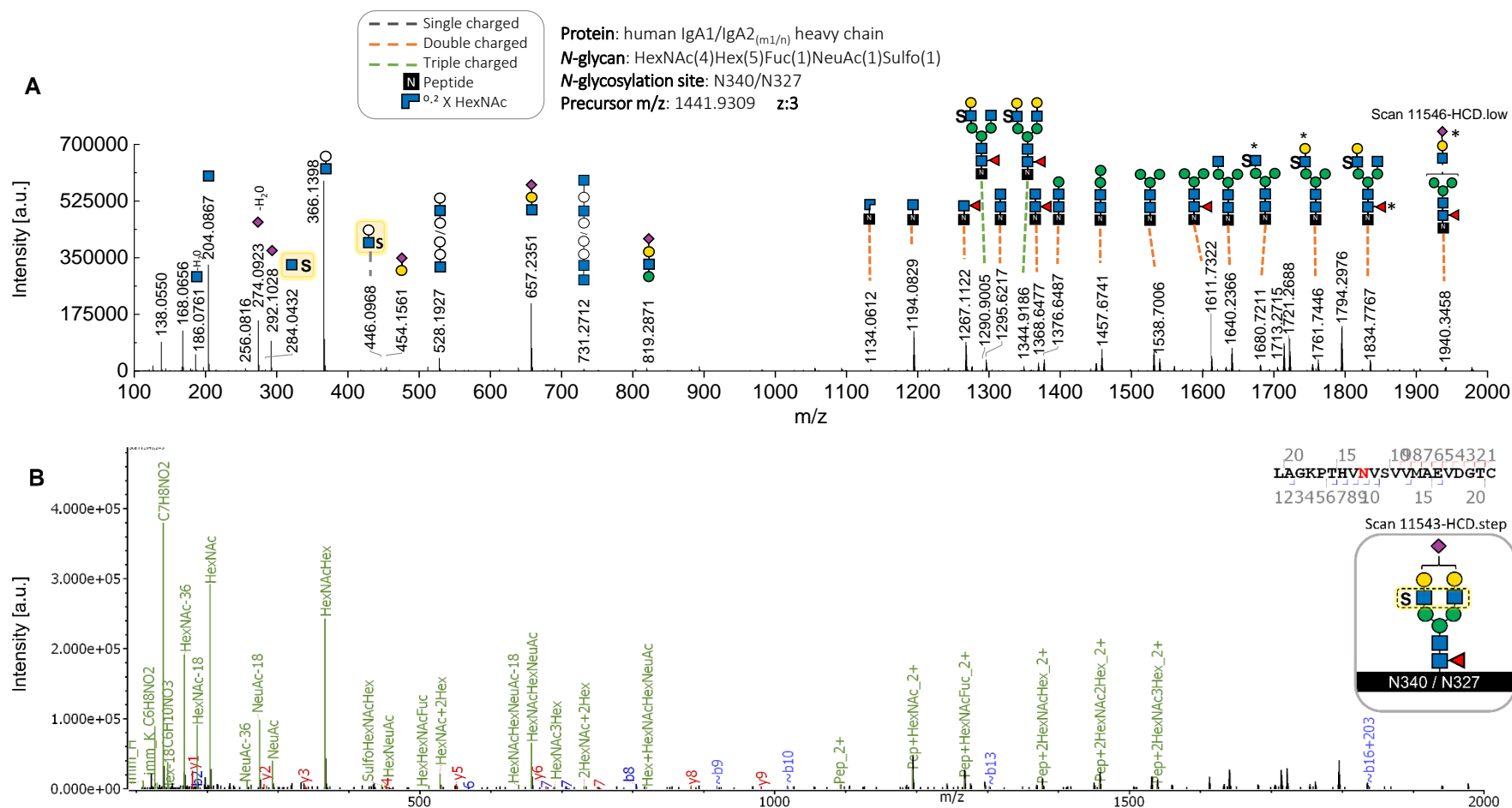
Supplementary Figure A.2.5. Annotated spectra of a glycopeptide bearing the sulfated N-glycan: HexNAc(4)Hex(5)Fuc(1)NeuAc(2)Sulfo(2). (A) Description of the N-glycan structure through HCD.low spectrum shows B and Y ions confirming one sulfation linked to antenna HexNAc, while the linkage of the second sulfation cannot be fully deduced and it can be associated to antenna HexNAc or Hex. The N-glycan icon with an asterisk symbol (*) depicts the N-glycan composition by a Y ion assigned to the second isotope peak of the expected m/z. (B) HCD.step spectrum annotated by Byonic™ recognizes b and y ions corresponding to an amino acid sequence of this N-glycopeptide containing the homologous site N340-IgA1/N327-IgA2_(m1/n). The HCD.low spectrum shows Sulfo₁HexNAc₁/284.04367 [M+H]⁺ oxonium ion, confirming HexNAc sulfation. The HCD.low and HCD.step spectra that most extensively describe the glycoform and peptide sequence, acquired within the shortest time, were selected for representing this glycopeptide.



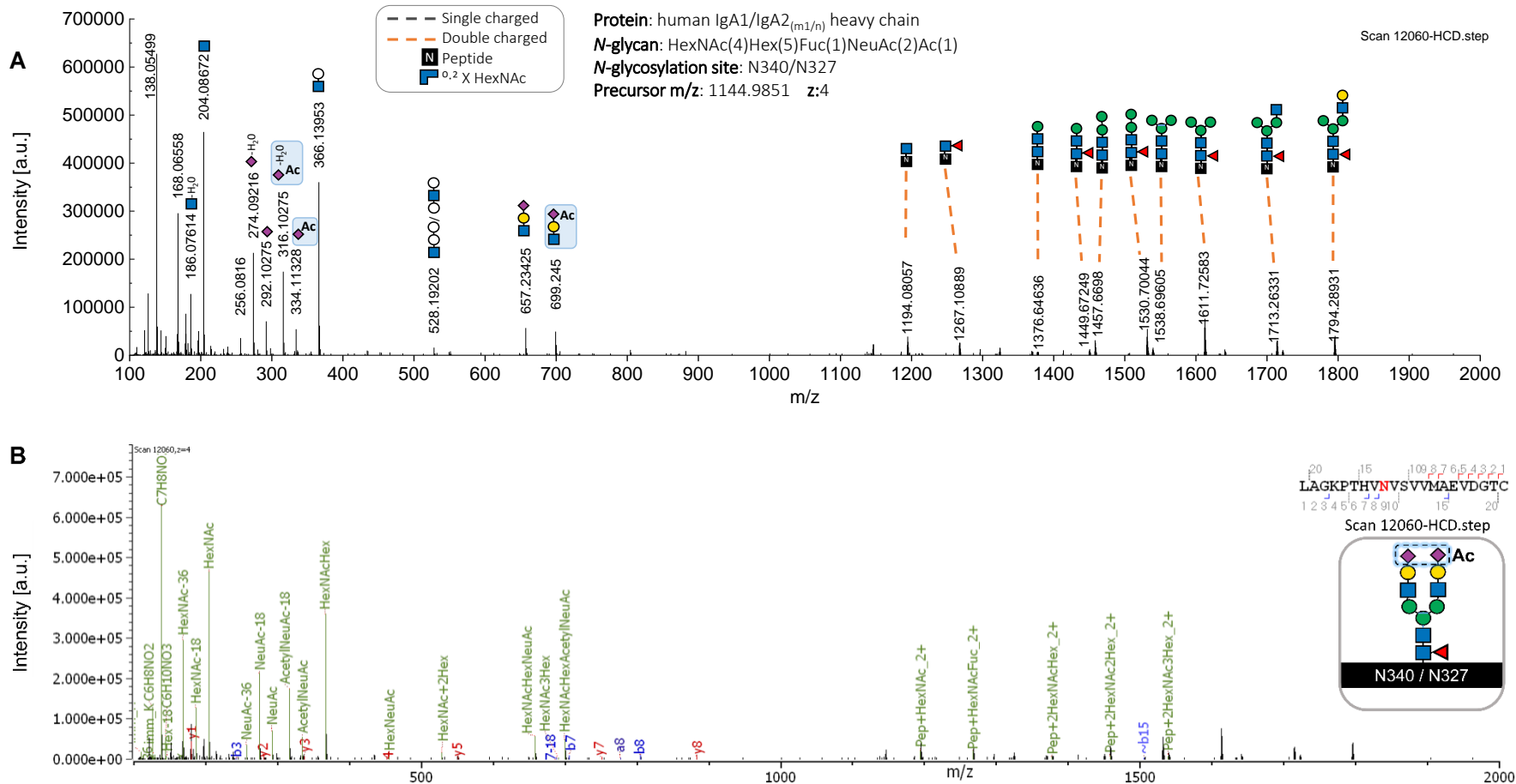
Supplementary Figure A.2.6. Annotated data of a glycopeptide bearing the sulfated N-glycan: HexNAc(5)Hex(5)Fuc(1)NeuAc(2)Sulfo(1). (A) Description of bisected N-glycan structure through HCD.low spectrum shows Y ions indicating sulfation most likely in one antenna HexNAc and not in the bisecting HexNAc. (B) The Byonic™ annotation of the HCD.step spectrum describes that the amino acid sequence of this N-glycopeptide bearing the homologous site N340-IgA1/N327-IgA2_(m1/n). The HCD.step spectrum also shows the oxonium ion confirming HexNAc sulfation (Sulfo₁HexNAc₁/284.0439 [M+H]⁺). The HCD.low and HCD.step spectra that most extensively describe the glycoform and peptide sequence, acquired within the shortest time, were selected for representing this glycopeptide.



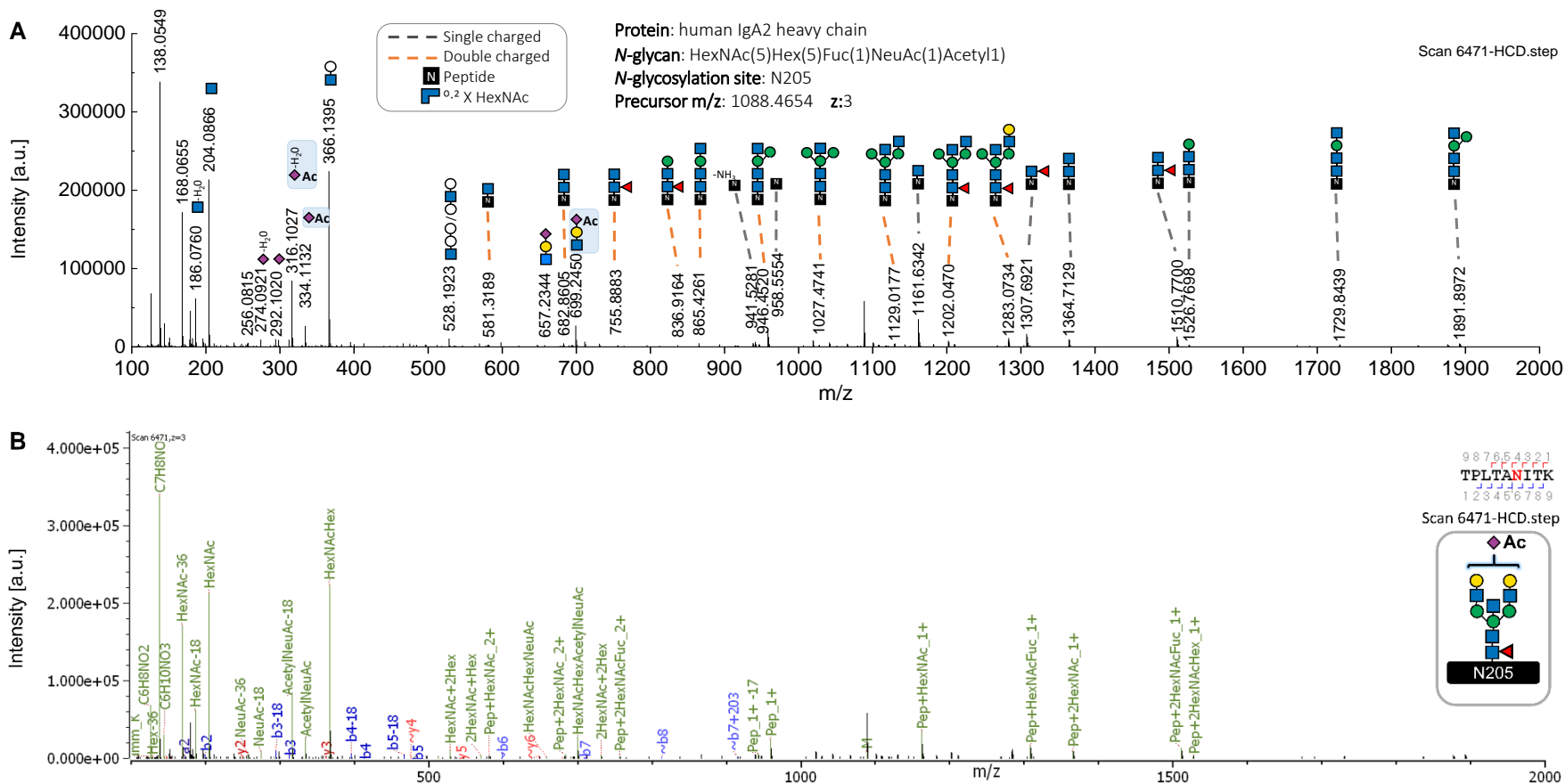
Supplementary Figure A.2.7. Annotated spectra of a glycopeptide bearing the sulfated N-glycan: HexNAc(4)Hex(6)Fuc(2)NeuAc(1)Sulfo(1). (A) Description of di-fucosylated hybrid N-glycan structure through HCD.low spectrum shows B and Y ions indicating HexNAc sulfation. The N-glycan icon with an asterisk symbol (*) depicts the N-glycan composition with a Y ion assigned to the second isotope peak of the expected m/z. (B) Byonic™ annotation of the HCD.step spectrum shows b and y fragment ions of a peptide containing homologous site N340-IgA1/N327-IgA2_(m1/n). The HCD.step spectrum shows the oxonium marker ions confirming HexNAc sulfation (Sulfo₁HexNAc₁/284.0433 [M+H]⁺). The HCD.low and HCD.step spectra that most extensively describe the glycoform and peptide sequence, acquired within the shortest time, were selected for representing this glycopeptide.



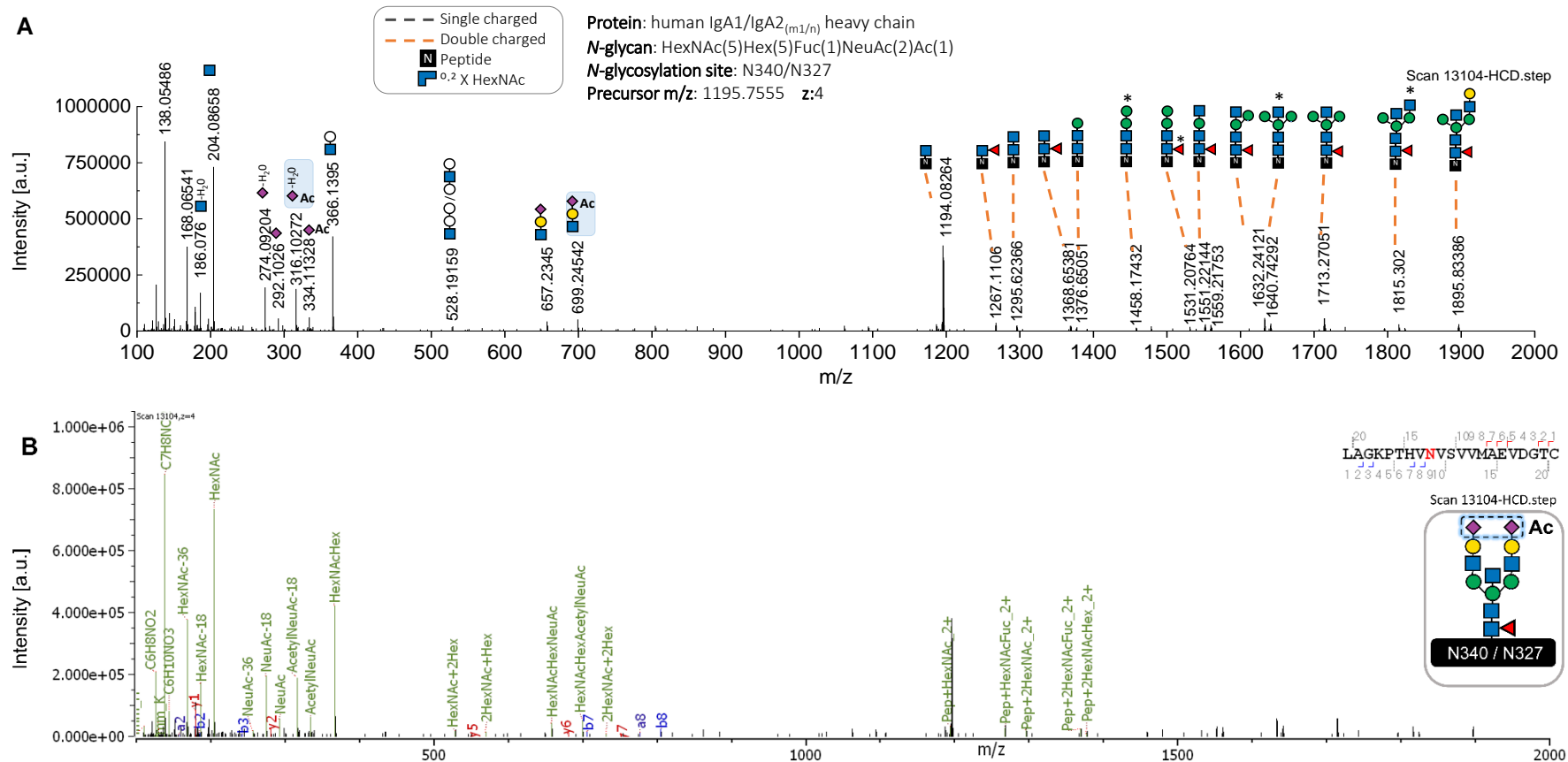
Supplementary Figure A.2.8. Annotated spectra of a glycopeptide bearing the sulfated N-glycan: HexNAc(4)Hex(5)Fuc(1)NeuAc(1)Sulfo(1). (A) Description of the N-glycan structure through HCD.low spectrum shows B and Y ions confirming one sulfation linked to antenna HexNAc. The N-glycan icon with an asterisk symbol (*) depicts the N-glycan composition by a Y ion assigned to the second isotope peak of the expected m/z. (B) HCD.step spectrum annotated by Byonic™ recognizes b and y ions corresponding to an amino acid sequence of this N-glycopeptide containing the homologous site N340-IgA1/N327-IgA2_(m1/n). The HCD.step spectrum shows the oxonium marker ion Sulfo₁HexNAc₁Hex₁/446.0967 [M+H]⁺ and also contains the oxonium ion Sulfo₁HexNAc₁/284.0437 [M+H]⁺ (the label of the last ion is not displayed in the plot). The HCD.low and HCD.step spectra that most extensively describe the glycoform and peptide sequence were selected for representing this glycopeptide.



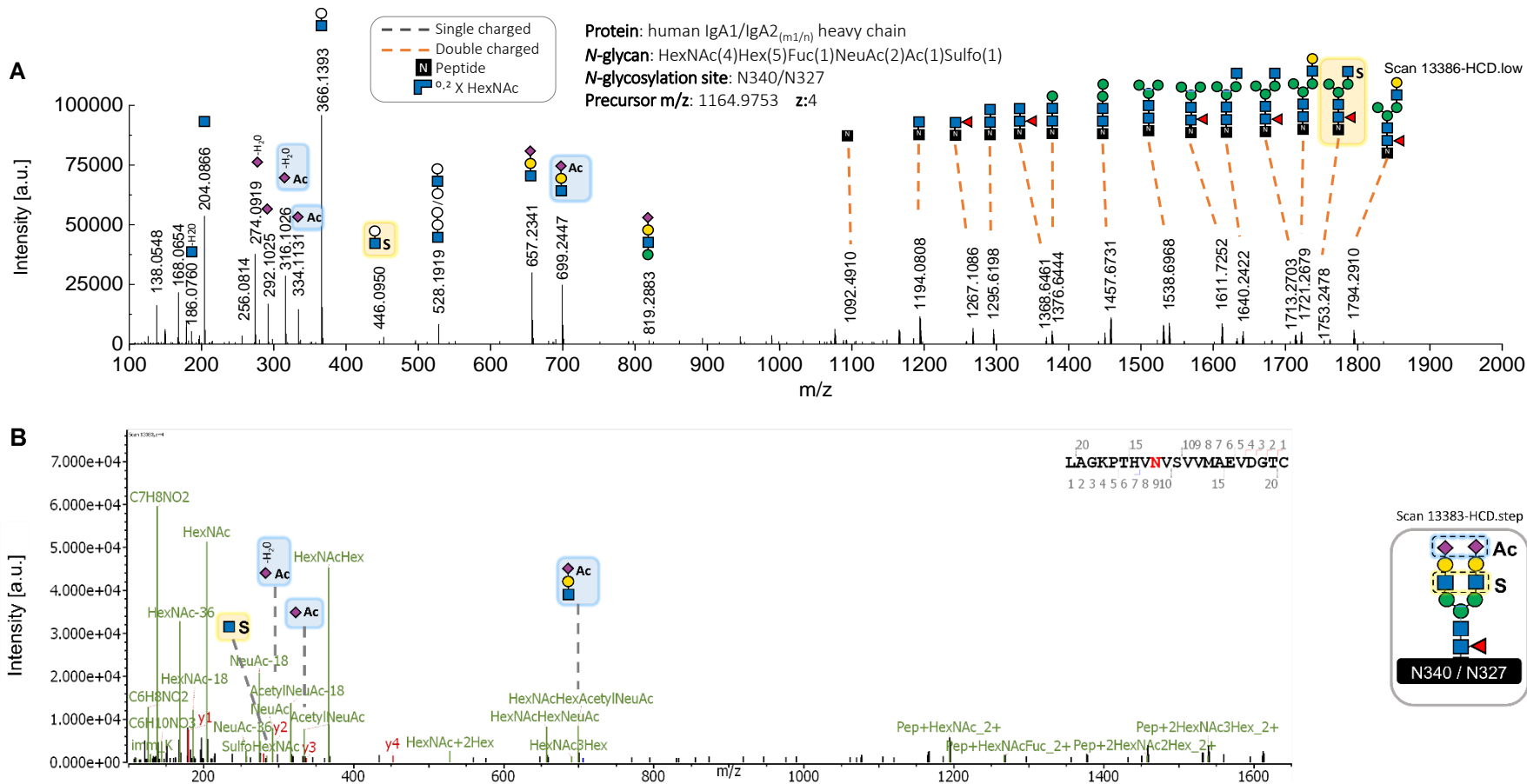
Supplementary Figure A.2.9. Annotated data of a glycopeptide containing NeuAc O-acetylation in the N-glycan HexNAc(4)Hex(5)Fuc(1)NeuAc(2)Ac(1). (A) Description of N-glycan structure through HCD. low spectrum shows B and Y ions providing structural evidence on NeuAc O-acetylation (highlighted in blue). (B) Byonic™ annotation of the HCD.step spectrum identifies b and y ions released from a peptide containing the homologous site N340-IgA1/N327-IgA2_(m1/n). Oxonium marker ions confirming NeuAc O-acetylation (Acetyl₁NeuAc₁/316.1027 [M-H₂O+H]⁺, Acetyl₁NeuAc₁/334.1133 [M+H]⁺ and HexNAc₁Hex₁Acetyl₁NeuAc₁/699.2454 [M+H]⁺).



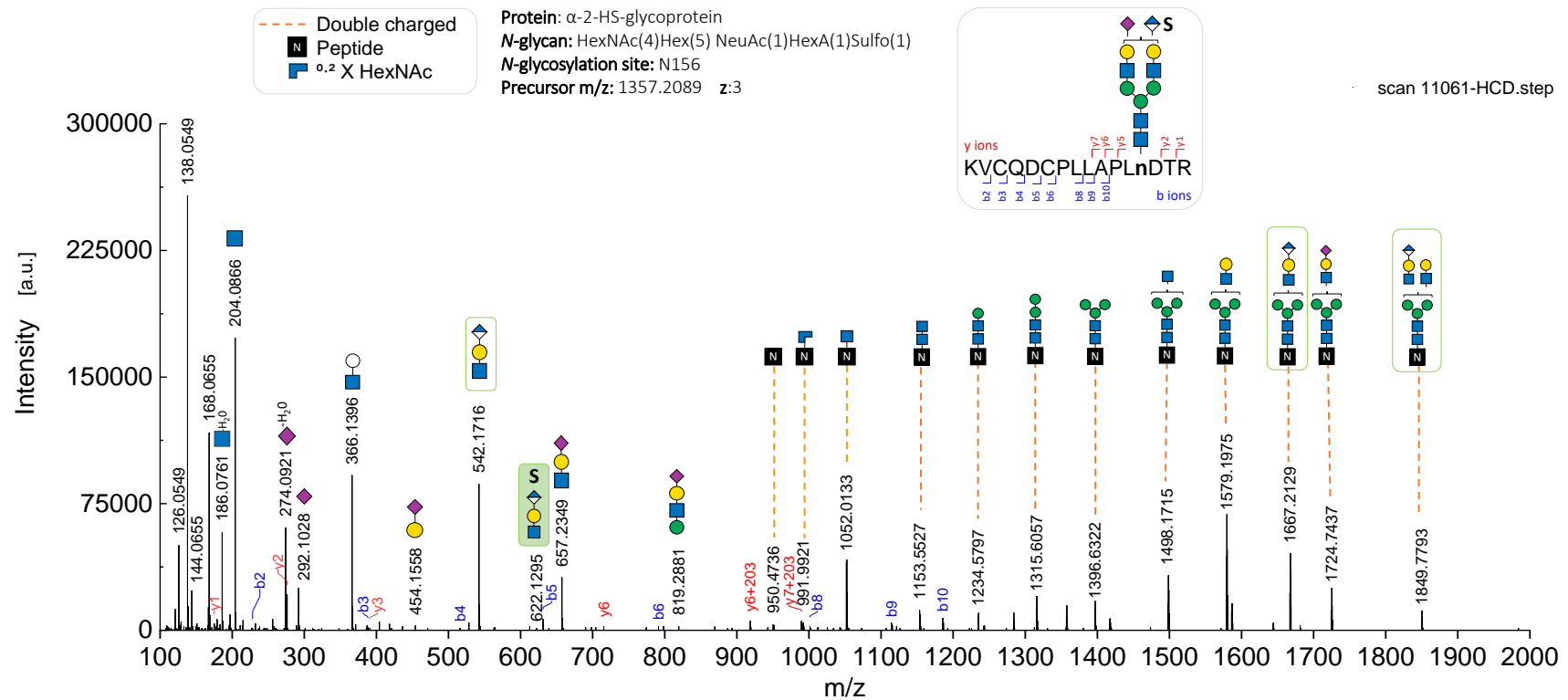
Supplementary Figure A.2.10. Annotated spectra of a glycopeptide containing the NeuAc O-acetylated N-glycan: HexNAc(5)Hex(5)Fuc(1)NeuAc(1)Ac(1). (A) Description of bisected N-glycan structure through HCD.step spectrum shows B ions indicating NeuAc O-acetylation (highlighted in blue). (B) Byonic™ annotation of the HCD.step spectrum displays b and y ions that identify an N-glycopeptide containing the site N205-IgA2_(m1/m2/n). Oxonium marker ions confirming NeuAc O-acetylation (Acetyl₁NeuAc₁/316.1027 [M+H-H₂O]⁺, Acetyl₁NeuAc₁/334.1132 [M+H]⁺ and HexNAc₁Hex₁Acetyl₁NeuAc₁/699.2450 [M+H]⁺).



Supplementary Figure A.2.11. Annotated data of glycopeptide containing NeuAc O-acetylation in the N-glycan HexNAc(5)Hex(5)Fuc(1)NeuAc(2)Ac(1). (A) Description of bisected N-glycan structure through HCD.step spectrum shows B ions indicating NeuAc O-acetylation (highlighted in blue). The N-glycan icon with an asterisk (*) depicts the N-glycan composition by a Y ion assigned to the second isotope peak of the expected ion m/z. (B) Byonic™ annotation of the HCD.step spectrum identifies fragment ions released from an N-glycopeptide containing the homologous site N340-IgA1/N327-IgA2_(m1/n). Oxonium marker ions confirming NeuAc O-acetylation (Acetyl₁NeuAc₁/316.1027 [M-H₂O+H]⁺, Acetyl₁NeuAc₁/334.1133 [M+H]⁺ and HexNAc₁Hex₁Acetyl₁NeuAc₁/699.2454 [M+H]⁺).



Supplementary Figure A.2.12. Annotated data of a glycopeptide containing NeuAc O-acetylation plus HexNAc sulfation in the N-glycan: HexNAc(4)Hex(5)Fuc(1)NeuAc(2)Ac(1)Sulfo(1). (A) Description of N-glycan structure through HCD.low spectrum shows B and Y ions providing structural evidence on NeuAc O-acetylation (highlighted in blue) and antenna sulfation (highlighted in yellow). (B) Byonic™ annotation of the HCD.step spectrum identifies fragment ions released from an N-glycopeptide containing the homologous site N340-IgA1/N327-IgA2_(m1/n). The HCD.step spectrum also contains oxonium marker ions confirming HexNAc sulfation (Sulfo₁HexNAc₁/284.0431 [M+H]⁺, highlighted in yellow) and NeuAc O-acetylation (Acetyl₁NeuAc₁/316.1027 [M-H₂O+H]⁺, Acetyl₁NeuAc₁/334.1133 [M+H]⁺ and HexNAc₁Hex₁Acetyl₁NeuAc₁/699.2454 [M+H]⁺).



Supplementary Figure A.2.13. Annotated data of a glycopeptide potentially containing HexA sulfation in the N-glycan HexNAc(4)Hex(5)NeuAc(1)HexA(1)Sulfo(1) of a contaminant protein. Annotation of N-glycan structure through HCD.step spectrum shows B and Y ions providing structural evidence on HexA as one antenna capping sugar (fragment ions framed in green color). The oxonium marker ion confirming HexA sulfation (HexNAc1Hex1HexA1Sulfo1/622.1295 [M+H]⁺, highlighted in green) resembles the glycopeptide human natural killer-1 (HNK-1) composed of glucuronic acid. The b and y fragment ions (in blue and red color, respectively) demonstrate a tryptic peptide bearing the site N156 in α -2-HS-glycoprotein. The N-glycopeptide was identified through Byonic™.

Declaration of honor

I hereby declare that I produced this thesis without prohibited external assistance and that none other than the listed references and tools have been used.

In the case of co-authorship, especially in the context of a cumulative dissertation, the own contribution is correctly and completely stated. I did not make use of any commercial consultant concerning graduation. A third party did not receive any nonmonetary perquisites neither directly nor indirectly for activities which are connected with the contents of the presented thesis. All sources of information are clearly marked, including my own publications.

In particular I have not consciously:

- Fabricated data or rejected undesired results
- Misused statistical methods with the aim of drawing other conclusions than those warranted by the available data
- Plagiarized data or publications
- Presented the results of other researchers in a distorted way

I do know that violations of copyright may lead to injunction and damage claims of the author and also to prosecution by the law enforcement authorities.

I hereby agree that the thesis may need to be reviewed with an electronic data processing for plagiarism.

This work has not yet been submitted as a doctoral thesis in the same or a similar form in Germany or in any other country. It has not yet been published as a whole.

Magdeburg,

Frania Jaqueline Zúñiga-Bañuelos

List of publications

Publications of this dissertation

First Manuscript:

[Preprint] **Zuniga-Banuelos, F. J.**, Hoffmann, M., Reichl, U., Rapp, E. (2023). New avenues for human blood plasma biomarker discovery via improved in-depth analysis of the low-abundant *N*-glycoproteome. bioRxiv. doi.org/10.1101/2023.10.22.562384.

[Peer-reviewed article] **Zuniga-Banuelos, F. J.**, Hoffmann, M., Reichl, U., Rapp, E. (2025). New avenues for human blood plasma biomarker discovery via improved in-depth analysis of the low-abundant *N*-glycoproteome. Engineering. doi.org/10.1016/j.eng.2024.11.039.

Second Manuscript:

[Preprint] **Zuniga-Banuelos, F. J.**, Lemke, G., Hoffmann, M., Reichl, U., Rapp, E. (2024). Immunoglobulin A Carries Sulfated *N*-glycans Primarily at the Tailpiece Site – An Oxonium-Ion-Guided Approach for Site-Specific *N*-glycan Identification. bioRxiv. doi.org/10.1101/2024.06.06.597690.

[Peer-reviewed article] **Zuniga-Banuelos, F. J.**, Lemke, G., Hoffmann, M., Reichl, U., Rapp, E. (2025). Immunoglobulin A Carries Sulfated and *O*-acetylated *N*-glycans Primarily at the Tailpiece Site – Strategies for Site-Specific *N*-glycan Identification. Front. Mol. Biosci. doi: 10.3389/fmolb.2025.1595173.

Supervised theses

Edel, S. (2021). Exploration of experimental approaches for the estimation of protein and peptide concentration in unknown samples, prior to proteomic analysis. Bachelor Thesis, Otto-von-Guericke-Universität, Magdeburg.

Lemke, G. (2022). Investigation of *N*-glycan sulfation on human Immunoglobulin A. Bachelor Thesis, Otto-von-Guericke Universität, Magdeburg.

Talks

Zuniga-Banuelos, F. J., Lemke, G., Hoffmann, M., Reichl, U., Rapp, E. (2025). *Expanding Glycoproteomics: Including N-glycopeptides Bearing Glucuronidated, O-Acetylated or Sulfated N-glycans in the Analysis of Human Serum Proteins such as Immunoglobulin A*. Talk presented at GlycoBioTec 2025. Berlin (Germany). 2025-02-09- 2025-02-12.

Zuniga-Banuelos, F. J., Lemke, G., Hoffmann, M., Reichl, U., Rapp, E. (2024). *Discovering Rare Sialylated N-glycans Holding O-Acetylation or Sulfation within the Site-Specific N-glycosylation of Human Serum IgA*. Talk presented at SialoGlyco 2024 (International Conference in SialoGlycobiology 2024). Lille (France). 2024-06-04 - 2024-06-07.

Zuniga-Banuelos, F. J., Lemke, G., Hoffmann, M., Reichl, U., Rapp, E. (2024). *Including N-glycopeptides Bearing Glucuronidated, O-Acetylated or Sulfated N-glycans in the Analysis of Human Serum Proteins such as Immunoglobulin A*. Talk presented at 33rd Joint Glycobiology Meeting (JGM). Hannover (Germany). 2024-09-29 - 2024-10-01.

Zuniga-Banuelos, F. J., Lemke, G., Hoffmann, M., Reichl, U., Rapp, E. (2024). *Expanding Glycoproteomics: Including N-glycopeptides Bearing Glucuronidated, O-Acetylated or Sulfated N-glycans in the Analysis of Human Serum Proteins such as Immunoglobulin A*. Talk presented at EGC 2024 webinar (European Glycoscience Community Webinar Series). online. 2024-09-26.

Zuniga-Banuelos, F. J., Hoffmann, M., Budimir, I., Reichl, U., Rapp, E. (2023). *New avenues for biomarker discovery in human blood plasma via an improved in-depth analysis of the low-abundant N-glycoproteome*. Talk presented at GlycoBioTec 2023. Berlin, Germany. 2023-01-17 - 2023-01-19.

Posters

Zuniga-Banuelos, F. J., Hoffmann, M., Reichl, U., Rapp, E. (2024). *In-depth N-glycoproteomic analysis reveals strategies to elucidate unanticipated N-glycans in the low-abundant human blood plasma N-glycoproteome*. Poster presented at 23rd Human Proteome Organization World Congress (HUPO World Congress), Dresden (Germany).

Zuniga-Banuelos, F. J., Hoffmann, M., Budimir, I., Reichl, U., Rapp, E. (2023). *New avenues for human blood plasma biomarker discovery via improved in-depth analysis of the low-abundant N-glycoproteome*. Poster presented at GlycoBioTec 2023, Berlin, Germany.

Lemke, G., **Zuniga-Banuelos, F. J.,** Grote, V., Hoffmann, M., Rapp, E., Reichl, U. (2023). *N-glycan Sulfation on Human IgA: In-Depth N-glycoproteomic Analyses of Sulfated N-glycan Site-Specificity and Distribution on Human IgA Isotypes*. Poster presented at GlycoBioTec 2023, Berlin, Germany.

Zuniga-Banuelos, F.J., Hoffmann, M., Reichl, U., Rapp, E. (2019). *In-Depth N-Glycoproteomic Analysis of Human Blood Plasma Proteins*. Poster presented at GlycoBioTec 2019, Berlin, Germany.

Datasets

New avenues for human blood plasma biomarker discovery via improved in-depth analysis of the low-abundant N-glycoproteome. The MS raw files were deposited to the ProteomeXChange Consortium, dataset PXD042039 (<https://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD042039>), through MassIVE repository (dataset identifier MSV000091870).

Immunoglobulin A Carries Sulfated and O-acetylated N-glycans Primarily at the Tailpiece Site. The MS raw files were deposited to the ProteomeXChange Consortium, identifier PXD060281 (<https://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD060281>), through MassIVE repository (dataset identifier MSV000096980).