

The Mass Spectrometry Toolkit for Glycoprotein Characterisation: Development of Novel Analytical Methods and Technologies for Glycomics and Glycoproteomics

Inaugural-Dissertation
to obtain the academic degree
Doctor rerum naturalium (Dr. rer. nat.)

Submitted to the Department of Biology, Chemistry and Pharmacy
of Freie Universität Berlin

by

Kathirvel Alagesan

from Madurai, Tamil Nadu, India

2016

This work was performed between Oct-2012 and Nov-2016 under the guidance of Dr. Daniel Kolarich in the Department of Biomolecular Systems, Max Planck Institute of Colloids and Interfaces Potsdam, and the Institute of Chemistry and Biochemistry, Freie Universität Berlin.

1st reviewer: Dr. Daniel Kolarich

2nd reviewer: Prof. Dr. Christian Freund

Date of oral defense: 20-12-2016

ACKNOWLEDGEMENTS

Foremost, I am deeply grateful to my supervisor, Dr. Daniel Kolarich, Group leader, Glycoproteomics group, Department of Biomolecular Systems, Max Planck Institute of Colloids and Interfaces for his patience, motivation, detailed & constructive comments and for his constant support throughout this work.

Besides my supervisor, I would like to thank Dr. Daniel Varón Silva, Group leader, GPI and Glycoproteins, Department of Biomolecular Systems, Max Planck Institute of Colloids and Interfaces for his valuable advices.

I would like to express my sincere gratitude to Prof. Dr. Peter H. Seeberger for giving me this wonderful opportunity to carry out my PhD thesis at Department of Biomolecular Systems, Max Planck Institute of Colloids and Interfaces, Berlin, Germany.

Additionally, I would like to thank my thesis committee advisor Prof. Dr. Christian Freund, Institute of Chemistry and Biochemistry, Freie Universität, Berlin, Germany.

I am grateful to Beilstein Institute for their financial support and especially to Dr. Carsten Kettner und Christina Keil for their constant support and help. Also I would like to Thank Maurice Grube, fellow scholarship recipient for being a great travel buddy.

In addition, I would like to thank all my collaboration with the department Dr. Kerry Gilmore, Dr. Chakkumkal Anish, Bopanna Monnanda Ponnappa and also Dr. Paula M Mendes, University of Birmingham; Dr. Rob Field & Ellis O'Neil, John Innes Center; Friedrich Altmann, BOKU; Nicki Packer & Matthew Campbell, Macquarie University, Australia.

I thank all the members of my group, Uwe Möglinger, Hannes Hinneburg, Kathrin Stavenhagen, Johanna Knaack, Andreia Almeida, Falko Schirmeister for their constant support throughout the project and for the pleasant working atmosphere. Besides my group, I would also like to thank the members of the GPI and Glycoproteomics group especially Dana Michel for teaching me patiently the tricks of peptide chemistry.

I would also like to thank all my wonderful colleagues and friends in the BMS department Matthew Plutschack, Stella Vukelic, Bopanna P. Monnanda, Dr. Adam Calow, Dr. Martina Delbianco; Monika Garg, Ankita Malik, Pietro Dallabernardina, Silvia Varela, Dr. Bartholomäus Pieber, Dr. Maria Antonietta Carillo, Dr. Benjamin Schumann, Dr. Sharavathi G. Parameswarappa, Dr. Marilda P. Lisboa, Dr. Madhu Emmadi, Feifei Xu, Priya Bharate, Dr. Claney.

Huge thanks to Malte Gölden for being there being there for me and putting up with me and also to Gölden family to consider me as one of theirs. Most importantly of all, I would like to thank my parents and brothers for all their support and love.

கற்றது கைமண் அளவு, கல்லாதது உலகளவு!!

"What you have learned is a mere handful; what you haven't learned is the size of the world"

- **Avvaiyar**

Summary

Glycosylation is one of the most complex yet common PTMs, which drastically enhances the functional diversity of proteins and influences their biological activity. Understanding relationships between structure, location, and function of glycoproteins requires detailed information about peptide sequence, glycan composition and sites of glycosylation. Tandem mass spectrometry (MS/MS) is a powerful tool to determine such data due to its ability to derive structural information about glycans and peptides as well as to localise sites of glycosylation. In this work several novel analytical and quantitative mass spectrometric tools have been developed for glycomics and glycoproteomics research.

Glycan identification and characterisation is crucial to correlate glycan structure to biological function. An in-depth glycan sequencing tool was developed that is based on Porous Graphitized Carbon Liquid Chromatography (PGC-LC) ElectroSpray Ionisation (ESI) tandem Mass Spectrometry (MS/MS). The diagnostic fragment ions observed in negative mode fragmentation in combination with the high separation capacity provided by the PGC was the basis to develop a reliable, fully customisable Glyco Relative retention time (RRT) MS/MS library that enables now high-throughput automated glycan identification and characterisation from PGC-LC ESIMS/MS data. Though this system works well for human glycans, glycomics from less well characterised species also requires knowledge on the monosaccharide building blocks. Many of these are indistinguishable by mass alone, and also detailed linkage information is not always obtained by MS/MS analyses. This obstacle was overcome by developing a novel, simple and sensitive method for unambiguous identification and linkage determination of various monosaccharides (including N-acetylneuraminic acids) using Reverse Phase (RP) – LC -ESI-MS/MS based approach that now can provide this information from minimal amounts of starting material.

Well-defined synthetic peptides and glycopeptides represent unique analytical glycoproteomics tools. Amino acid building blocks modified with large N-linked glycans for solid phase peptide synthesis (SPPS), however, are not commercially available. In the present work a fast and "environmentally friendly" method was developed for the fast and efficient isolation of a glycosylated asparagine building block carrying complex sialylated biantennary N-glycans from egg yolk. The

obtained glyco amino acid building blocks were applied in solid phase glycopeptide synthesis after converting them into Fmoc protected glyco-amino acids and selective protection of the sialic acids.

These building blocks were used to generate a synthetic glycopeptide library that fostered the systematic investigation and optimisation of various glycoproteomic sample preparation steps and analysis parameters. A novel, simplified technique for HILIC (hydrophilic interaction chromatography) based glycopeptide enrichment termed "Drop-HILIC" was developed and applied to systematically evaluate the mobile phase effect on ZIC-HILIC (zwitterionic type HILIC) glycopeptide enrichment. Acetonitrile, Methanol Ethanol and Isopropanol were tested, and Acetonitrile was found to provide the best compromise for the retention of both hydrophilic and hydrophobic glycopeptides. Though isopropanol enriched more glycopeptides, is also co-enriched more non-glycosylated peptides. Methanol was confirmed to be unsuitable for the purpose of ZIC-HILIC based glycopeptide enrichment.

This synthetic glycopeptide library also served as a basis for the systematic investigation of glycopeptide fragmentation and their ionisation behaviour in LC-MS-CID/ETD, which in turn allowed optimising analysis conditions for biologically relevant samples in glyco-biomarker research. An investigation on ETD fragmentation efficiency was performed on a panel of isobaric glycopeptides differing only in the position of glycosylation site, revealing that ETD fragmentation efficiencies depended on glycan size and glycosylation site location.

Severe limitations reported earlier for glycopeptide ionisation were overcome by the CaptiveSpray Nanobooster™ ionisation technique, which was systematically investigated using a commercially available Mass Spectrometry Retention Time calibration mix and selected synthetic glycopeptides. CaptiveSpray Nanobooster™ ionisation of glycopeptides depended on the physio-chemical properties of the dopant solvent, and acetonitrile provided the best results as it enhanced glycopeptide signal intensities and increased their charge state at the same time.

The in this thesis presented novel analytical and quantitative mass spectrometric tools provide a solid, sensitive and automatable basis that is now being applied to gain a better understanding of the biological role of protein glycosylation.

Zusammenfassung

Die Protein-Glykosylierung ist eine der komplexe und weit verbreitete, post-translationelle Modifikation, die auch einen wesentlichen Einfluss auf die Struktur und Aktivität von Proteinen hat. Um das Verhältnis zwischen Struktur, Position und Funktion der Glykoproteine zu verstehen sind detaillierte Daten zur Peptidsequenz, der Glykanstruktur und deren Position innerhalb des Glykoproteins essentiell. Die Tandem-Massenspektrometrie (MS/MS) stellt eine hierfür weit verbreite Methode dar um diese Daten zu akquirieren. Sie bietet die Möglichkeit, diese Strukturinformationen über Glykane und Peptide zu gewinnen. Darüber hinaus kann in bestimmten Fällen auch die Position der Glykane innerhalb eines Peptides bestimmt werden. Daher war es ein wesentliches Ziel dieser Arbeit neue analytische Methoden und quantitative massenspektrometrische Werkzeuge für die Glykomik und Glykoproteomik zu entwickeln.

Die Identifizierung von Glykanen und deren Charakterisierung ist essentiell um deren funktionelle Relevanz zu eruieren. Eine hochsensitive und selektive Glykomik-Methode wurde auf Basis der "Porous Graphitized Carbon" (PGC) Flüssigkeitschromatographie (LC), welche mit einer ElektroSpray Ionisation (ESI) MS/MS Detektion gekoppelt ist, entwickelt. Die Kombination aus der PGC-LC mit der Fragmentierung von negativ geladenen Produkt-Ionen ermöglicht das diagnostische Fragmentierungsmuster erhalten werden, welche wiederum eine exakte Identifizierung von Glykanstrukturen ermöglicht. Dies war die Basis für die Entwicklung einer umfassenden MS/MS Spektren-Bibliothek. Darüber hinaus wurde für all diese Glykanstrukturen auch eine "Relative Retention Time" (RRT) Bibliothek entwickelt, welche zusammen mit der MS/MS Spektren- Bibliothek eine noch genauere Strukturbestimmung der Glykane ermöglichte und gleichzeitig auch einen wesentlichen Schritt vorwärts zur Automatisierung der PGC-LC ESI MS/MS basierten Glykomik darstellte. Obwohl dieses System eine detaillierte Sequenzierung des humanen N- und O-glykoms ermöglichte, benötigt die Glykomanalytik aber auch das Wissen über die Identität der einzelnen Monosaccharide, aus welchen Glykane aufgebaut sind. Dies ist insbesondere wichtig wenn Glykoproteine nicht humanen Ursprungs charakterisiert werden sollen, da viele dieser Zuckermoleküle nicht einfach durch rein

massenspektrometrische Methoden eindeutig identifiziert werden können. Darüber hinaus ist es auch wichtig die Verbindungen zwischen den jeweiligen Monosacchariden zu kennen. Daher wurde im Rahmen dieser Arbeit auch neue, einfache und sensitive LC-MS basierte Methode für die eindeutige Identifizierung von Monosacchariden inklusive Sialinsäuren und deren Verknüpfungen entwickelt, welche es ermöglicht diese Informationen auch aus geringsten Probemengen zu bestimmen.

Definierte, synthetische Peptide und Glykopeptide sind einzigartige, analytische Werkzeuge für die Glykoproteomik. Aminosäure Bausteine, welche in der Festphasenpeptidsynthese verwendet werden können und auch komplexe, hochmolekulare N-Glykane tragen, sind aber nicht kommerziell verfügbar. Im Rahmen dieser Arbeit wurde eine schnelle und umweltfreundliche Methode entwickelt die eine einfache Isolierung von glykosylierten Asparagin-Bausteinen aus Hühnereigelb möglich machte. Die gewonnenen Aminosäurebausteine wurden nach entsprechender Modifizierung durch Fmoc und gezielter Einführung von Schutzgruppen für die Sialinsäuren in weiterer Folge in die Glykopeptid-Festphasensynthese eingesetzt und eine Bibliothek definierter synthetischer Glykopeptide wurde erstellt.

Diese Glykopeptid-Bibliothek wurde in weiterer Folge zur systematischen Evaluierung und Optimierung von unterschiedlichen, essentiellen Probenvorbereitungs- und Analysemethoden verwendet. Eine neue, vereinfachte Methode, "Drop HILIC", wurde für die HILIC ("hydrophilic interaction chromatography") basierte Anreicherung von Glykopeptiden entwickelt. Mit Hilfe dieser Methode war es nun möglich den Effekt der mobilen Phase auf die Glykopeptid-Anreicherung systematisch zu evaluieren. Acetonitril, Methanol, Ethanol und Isopropanol wurden hierfür getestet. Bei diesen Analysen zeigte sich, dass Acetonitril den besten Kompromiss aus selektiver Anreicherung und geringer Co-Anreicherung von nicht-Glykopeptiden lieferte. Obwohl durch die Verwendung von Isopropanol eine größere Anzahl von Glykopeptiden aus humanen Serum angereichert werden konnte, war auch der Anteil an unspezifisch angereicherten, nicht glykosylierten Peptiden wesentlich höher verglichen mit Acetonitril. Methanol hingegen war nicht geeignet für den ZIC-HILIC basierte Glykopeptidanreicherung.

Mit Hilfe der synthetischen Glykopeptid-Bibliothek konnte auch eine systematische Evaluierung des Glykopeptid-Ionisationsverhaltens und der ETD relevanten Fragmentierungseigenschaften untersucht werden. Dies erlaubte in weiterer Folge eine Verbesserung und Optimierung der Glykopeptidanalytik für biologisch relevante Fragestellungen im Bereich der Glyko-Biomarker Forschung. Durch diese Arbeiten zeigte sich, dass die ETD Fragmentierungs-Effizienz stark von der Glykangröße und der Position innerhalb eines Glykopeptides zusammenhängt.

Die "CaptiveSpray Nanobooster™" Ionisierungstechnik erlaubte es die schwachen Ionisierungseigenschaften von Glykopeptiden signifikant zu verbessern, was auch mittels eines kommerziell erhältlichen Retentionszeit-Kalibrierungsstandards und einer Reihe von synthetischen Glykopeptiden nachgewiesen wurde. Es zeigte sich darüber hinaus, dass die "CaptiveSpray Nanobooster™" Ionisierungstechnik stark von dem Lösungsmittelzusatz beeinflusst wird. Eine systematische Evaluierung unterschiedlicher Lösungsmittel zeigte, dass Acetonitril die besten Ergebnisse in Bezug auf die Glykopeptid-Signalstärke und die Protonierungseigenschaften von Glykopeptiden brachte.

Die im Rahmen dieser Arbeit neu entwickelten, analytischen und quantitativen Methoden zur Glykoproteomik bilden nun eine solide, sensitive und automatisierbare Plattform. Diese findet nun Anwendung um die funktionelle Rolle der Protein-Glykosylierung anhand individueller Glykoproteine besser untersuchen zu können.

Table of Contents

1. INTRODUCTION	1
1.1. Carbohydrates – undisputed molecular frontier of the cells	1
1.1.1. Carbohydrates of protein glycosylation	3
1.1.2. Biosynthesis	3
1.1.3. Biological Significance of protein glycosylation - all theories are correct!	6
1.2. MS based Glycoprotein analysis	9
1.3. Glycomics	10
1.3.1. Global glycomics	11
1.3.2. Tandem Mass Spectrometry provides deeper structural insights	12
1.3.3. Glyco-Bioinformatics – the next frontier to conquer in PGC based glycomics	14
1.3.4. Monosaccharide Linkage analysis	16
1.4. Glycoproteomics	17
1.4.1. Glycopeptide enrichment strategies in glycoproteomics	19
1.4.2. MS based Glycopeptide Characterisation	20
1.5. State of the Art glycopeptide synthesis	21
1.6. Objective of my thesis	25
1.7. References	26
2. IN-DEPTH STRUCTURAL CHARACTERISATION OF N- GLYCANS USING PGC-NANO-LC-ESI-MS/MS	34
2.1. Background	35
2.2. Materials And Methods	37
2.2.1. Glycan standards	37
2.2.2. SDS-PAGE and electro-blotting	41
2.2.3. Glycan release	41
2.2.3. PGC clean up	41
2.2.4. PGC nano LC-ESI IT-MS/MS analysis	42
2.2.5. spectral library design and implementation	44
2.2.6. Glycan identification and relative quantitation	45
2.3. Results And Discussion	46
2.3.1. Rationale and Library design	46
2.3.2. Negative ion mode PGC-LC-ESI-MS/MS glycomics	47
2.3.3. An Assessment of spectral quality and similarity scoring	49
2.3.4. Influence of The charge state on spectral alignment	52
2.3.5. Factors influencing glycan separation by PGC-LC	54
2.3.6. Better together: relative retention time + negative ion fragmentation provide two independent parameters for glycan structure elucidation	57
2.3.7. Influence of SPS parameter on quantification	69
2.3.8. quantitative global glycomics	73
2.4. Conclusion	78
2.5. Reference	80

3. DEVELOPMENT OF A NOVEL, ULTRASENSITIVE APPROACH FOR QUANTITATIVE CARBOHYDRATE COMPOSITION AND LINKAGE ANALYSIS	82
3.1. Introduction	82
3.2. Materials and methods	84
3.2.1. Reagents	84
3.2.2. General procedure for carbohydrate derivatisation	84
3.2.3. ESI-MS/MS and LC-ESI-MS/MS analysis of the derivatised monosaccharides	85
3.3. Results and Discussion	87
3.3.1. Rationale and Method development	87
3.3.2. Stereochemistry defines the fragmentation of 2-AB labelled, Permethyated monosaccharides	88
3.3.3. Chromatographic separation of 2-AB-ME labelled monosaccharides	92
3.3.4. Evaluation of the linkage analysis strategy using reference oligosaccharides	93
3.3.5. Method Application to N-glycans	97
3.3.6. Including sialic acids in the monosaccharide analysis equation	100
3.4. Optimised step-wise procedure for monosaccharide linkage and compositional analysis	104
3.5. Conclusion	106
3.6. References	108
4. DEVELOPMENT OF FAST AND EFFICIENT STRATEGIES FOR THE PRODUCTION OF GLYCOSYLATED AMINO ACID BUILDING BLOCK	110
4.1. Background	110
4.2. Materials and methods	111
4.2.1. Materials	111
4.2.2. Hexapeptide isolation	111
4.3.3. Pronase digestion	112
4.3.4. Glycopeptide synthesis	112
4.3.5. Glycopeptide purification	112
4.3.5. Exoglycosidase Digest	113
4.3. Results and discussion	113
4.3.1. Isolation of the glycosylated hexapeptide from fresh egg yolk	113
4.3.2. Synthesis of Fmoc protected glycosylated amino acid building blocks	120
4.3.3. Glycopeptide synthesis	122
4.4. Conclusion	129
4.4.1. Synthetic Glycopeptides are indispensable glycoproteomics tools	129
4.5. Reference	130

5. IT'S ALL ABOUT THE SOLVENT: ON THE IMPORTANCE OF THE MOBILE PHASE FOR ZIC -HILIC GLYCOPEPTIDE ENRICHMENT	132
5.1. Introduction	132
5.2. Materials And Methods	134
5.2.1. Materials	134
5.2.2. High abundance Serum Protein Depletion	134
5.2.3. In-solution Protease Digestion	134
5.2.4. HILIC Enrichment – micro spin	135
5.2.5. HILIC Enrichment – "Drop-HILIC"	135
5.2.6. Glycopeptide synthesis	136
5.2.7. LC-MS Analysis Parameters	136
5.3. Results And Discussion	138
5.3.1. Rationale and Experiment Design	138
5.3.2. Salt Removal is Crucial for Efficient HILIC Enrichment	139
5.3.3. Same, Same but Different: To Spin or to "Drop-HILIC"	140
5.3.4. Incubation Time Does Not Influence Drop-HILIC Enrichment	141
Efficiency	141
5.3.5. It's All About the Solvent – Influence of the Solvent System on ZIC-HILIC Glycopeptide Enrichment	142
5.3.6. Glycopeptide Enrichment from Human Serum using Drop-HILIC	147
5.4. Conclusion	149
5.5. References	150
6. TOWARDS UN-BIASED GLYCOPROTEOMICS - ENHANCING (GLYCO)PEPTIDE IONISATION USING THE CAPTIVESPRAY NANOBOOSTER™	152
6.1. Introduction	152
6.2. Materials And Methods	154
6.3. Results and Discussion	154
6.3.1. Rationale and study design	154
6.3.2. The quest for the optimal CaptiveSpray nanoBooster™ dopant	155
6.3.3. Performance characteristics of ETD glycopeptide fragmentation	163
6.4. Conclusion	170
6.5. Reference	172
7. APPENDIX	173
8. CURRICULUM VITAE	184

1. INTRODUCTION

1.1. CARBOHYDRATES – UNDISPUTED MOLECULAR FRONTIER OF THE CELLS

Every single eukaryotic cell is covered with a dense and complex array of glycans giving its characteristic bristle-brush like appearance (glycocalyx). These glycans are structurally diverse, incorporating a wide range of monosaccharide residues and glycosidic linkages. On the cell surface, glycans are usually found attached to proteins or lipids to form glycoconjugates such as glycoproteins or glycolipids (1). Most glycoconjugates are either secreted or span the plasma membrane with their glycans becoming the molecular frontier of the cell where they are involved in diverse biological processes such as cell-cell communication and recognition (2).

These glycoconjugates are differentiated based on their molecular characteristics. **Glycoproteins** can carry one or more glycans covalently attached to a polypeptide backbone usually via an amide linkage to an asparagine residue (*N*-linked) or an ether linkage to the side chain hydroxyl-group present in Serine or Threonine (mucin type *O*-linked) (1). In addition glycans can also be covalently attached to Tyrosine or Hydrolysine or Hydroxyproline residues (3). **Proteoglycans** are a class of glycoproteins with covalently attached **glycosaminoglycan (GAGs)** chains to serine/threonine residues of the core protein. In contrast to *N*-linked and *O*-linked glycans, glycosaminoglycans (GAGs) consist of linear disaccharide repeats that itself can also exist as free sugar chains. Glycoproteins can also be anchored to the lipid bilayer through a glycosylphosphatidylinositol (GPI) anchor to form **GPI-anchored proteins**, which themselves can also be *N*- and *O*-glycosylated. **Glycosphingolipids** are class of glycoconjugates where a glycan is linked to a lipid ceramide spanning the phospholipid bilayer (1) (Figure 1.2).

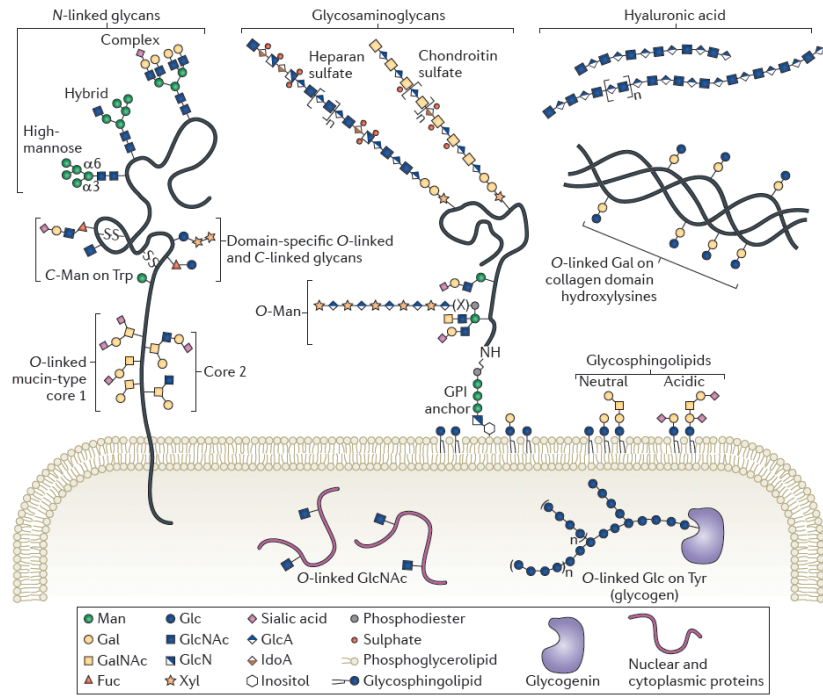


Figure 1.1: Different classes of glycans are attached to proteins and lipids. These glycoconjugates populate the surface of mammalian cells. Additionally, O-linked GlcNAc is found on many cytoplasmic and nuclear proteins (4).

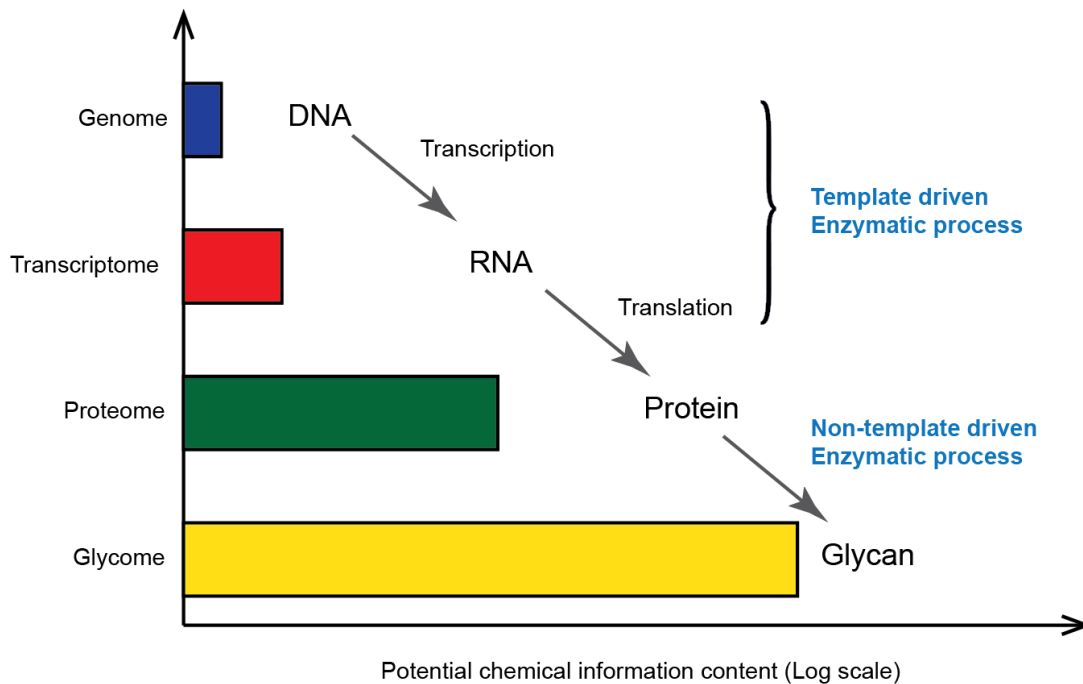


Figure 1.2: Protein glycosylation represents one of the major and most diverse protein post translation modifications. Glycosylation is a non-template based modification that has been shown to dramatically increase the functional diversity of proteins (5).

1.1.1. Carbohydrates of protein glycosylation

Eukaryotic cells produce a large diversity of biologically active proteins from a relatively small number of genes through a combination of alternative splicing events and numerous protein post translational modifications (PTM) (6, 7). As the name suggests, PTMs occur in a non-template based fashion after a protein has been synthesised from the initial DNA template (Figure 1.2). These PTMs are frequently used by nature to influence the biological activity of proteins and endow them with additional functions. Thus, PTMs are regarded as one of nature's tools for fine tuning protein function.

Protein glycosylation is one of the most widespread PTM found in eukaryotic organisms, particularly on secreted and outer membrane proteins where it is known to be crucially involved in processes such as cell-cell interaction and communication. The data available in protein databases suggests that as many as 70% of human proteins are in fact glycoproteins (8, 9).

1.1.2. Biosynthesis

Several amino acids can be the target for glycosylation within a protein sequence. To date 31 different types of sugar-amino acid linkages have been reported involving 8 amino acids and 13 different monosaccharides (10). Protein glycosylation is initiated in the Endoplasmic Reticulum (ER) and further continued in the Golgi apparatus. In contrast to DNA and protein biosynthesis, glycosylation is a non-template driven process involving different cell organelles and highly influenced by multiple endogenous and exogenous factors (11). It is a dynamic process mediated by highly synchronised cellular machinery involving various membrane-bound glycosyltransferases and glycosidases that are responsible for either adding or removing monosaccharide units, respectively (12-16).

The most common and best-studied types of protein linked glycans are the so called *N*- and *O*-glycans that are essentially present on all secreted and plasma membrane proteins. Understanding protein glycosylation also requires a basic knowledge on biosynthetic key events of the glycosylation pathway (11), which are briefly described in the following paragraphs.

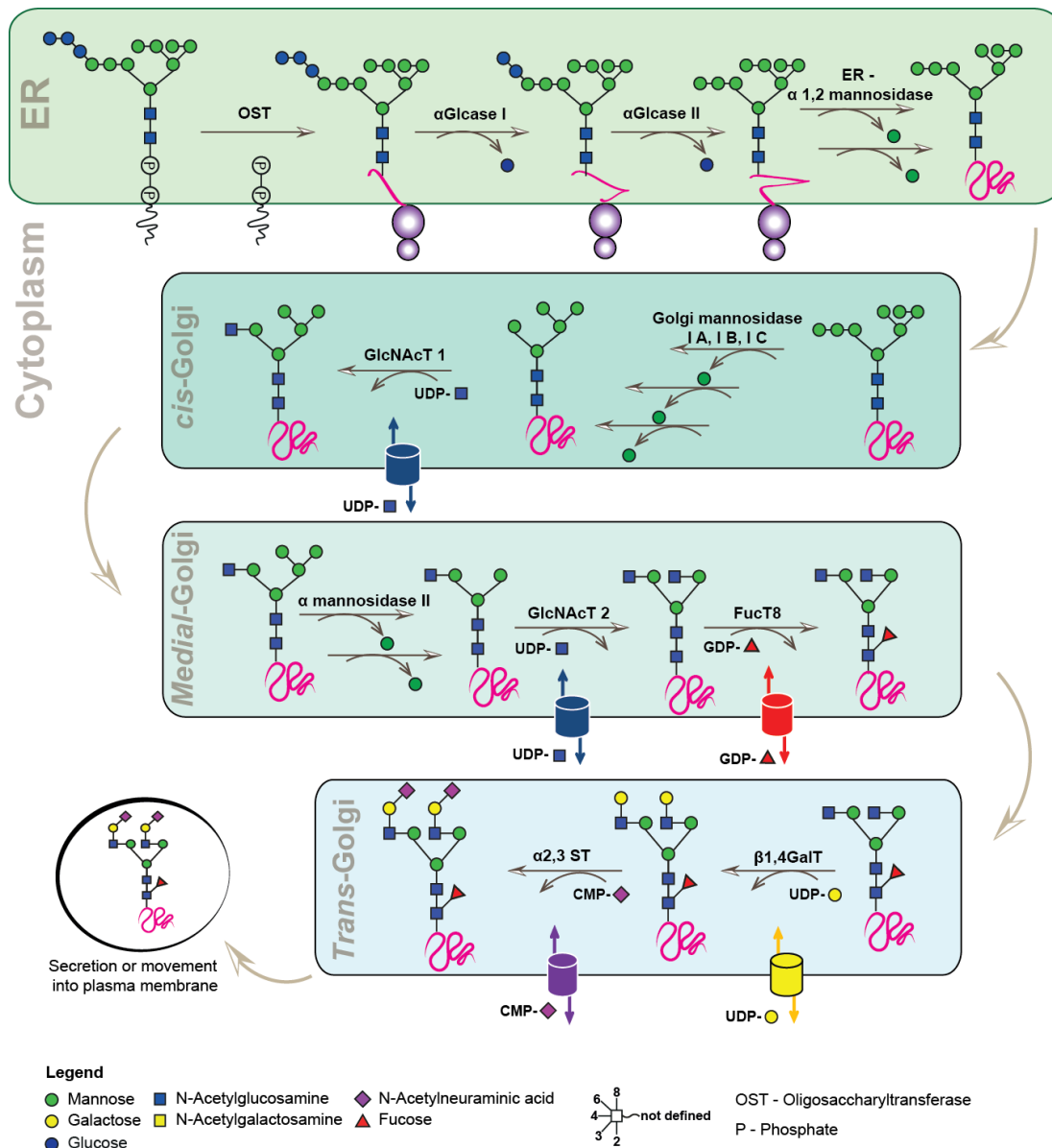


Figure 1.3: Biosynthesis of mammalian N-glycans (17)

A preassembled oligosaccharide comprising of two GlcNAc, nine Man and three Glc residues (left) is co-translationally transferred onto specific asparagine residues of a nascent protein chain. In the course of protein folding, this structure undergoes further processing by the sequential removal of three Glc residues, mediated by α -glucosidases (α Glcase) I and II, and four Man residues, mediated by α -mannosidases (α Mannosidase). Subsequently, a first antenna GlcNAc is added to the Man5GlcNAc2 structure by the action of the enzyme N-acetylglucosamine Transferase I (GlcNAcT1). Subsequently, two other Man residues are removed by α -mannosidase II (α Mannosidase II) and a second GlcNAc residue is added by GlcNAcT2. The resulting structure can be further elongated e.g. by the addition of galactose residues, mediated by members of the galactosyltransferase family such as β 1,4GalT and/or by core-linked fucose (mediated by FucT8) and/or bisecting GlcNAc (mediated by GlcNAcT3). The addition of these modifications is not mutually exclusive and cell type dependent. After the addition of galactose, the sugar chains are frequently found to be terminated by sialic acids which can be linked either via α 2,3 or α 2,6 to galactose.

Protein *N*-Glycosylation is the most complex, yet frequent co- and post translational modification initiated during protein translation within the Endoplasmic Reticulum (ER). *N*-glycosylation is initiated by an “en-bloc” transfer of the GlcNAc₂Man₉Glc₃ oligosaccharide that is preassembled on a dolichol-linked pyrophosphate donor embedded in the ER membrane. This 14mer oligosaccharide is transferred onto Asn residues present in the consensus motif Asn-Xxx-Ser/Thr/Cys (where Xxx is any amino acid except Pro) in the nascent polypeptide chain by the highly conserved oligosaccharyltransferase (OST) enzyme complex (Figure 1.3). Directly following its attachment this oligosaccharide plays an important role in protein folding where it acts as a signal for chaperons such as calnexin and calreticulin that control protein folding within the ER (18, 19). After correct protein folding these *N*-glycan chains undergo subsequent trimming by the action of ER-resident glucosidases and mannosidase and the glycoprotein is exported to the cis-Golgi for further processing. Subsequent glycan trimming followed by addition of particular monosaccharide units within the entire Golgi is catalysed by numerous specific glycosidases and glycosyltransferases that result in the formation of various glycoforms for each glycoprotein (20).

Unlike *N*-glycan synthesis, *O*-glycosylation is not known to occur on any consensus sequence motifs. It is also initiated on the already completely folded protein within the Golgi (21, 22). The first step is mediated by a family of specific polypeptide *N*-acetyl galactosaminyltransferases (ppGalNAcTs), which attach a GalNAc onto Ser/Thr residues of a protein (Figure 1.4). These initial GalNAcs can further be elongated by the subsequent action of specific transferases to yield several core structures. At present, there are 8 core *O*-glycan structures known, of which Core 1-4 represent the most predominant ones.

The glycan structures attached to proteins can be highly complex, as monosaccharides provide numerous possibilities how they can be attached to each other. Unlike linear nucleic acid or polypeptide structures glycan structures are not just defined by their particular sequence but also by branching and anomeric linkage between the individual monosaccharide building blocks. Enzymes involved in glycans biosynthesis are sensitive to subtle changes in the cellular environment, and glycans presented by a particular glycoprotein can also be a reflection of the physiological state of a cell (23, 24). As a result,

glycoproteins typically occur as a complex mixture, which can be either due to the presence of structurally different glycans at a single site of glycosylation (micro-heterogeneity) or due to site occupancy (macro-heterogeneity) (16).

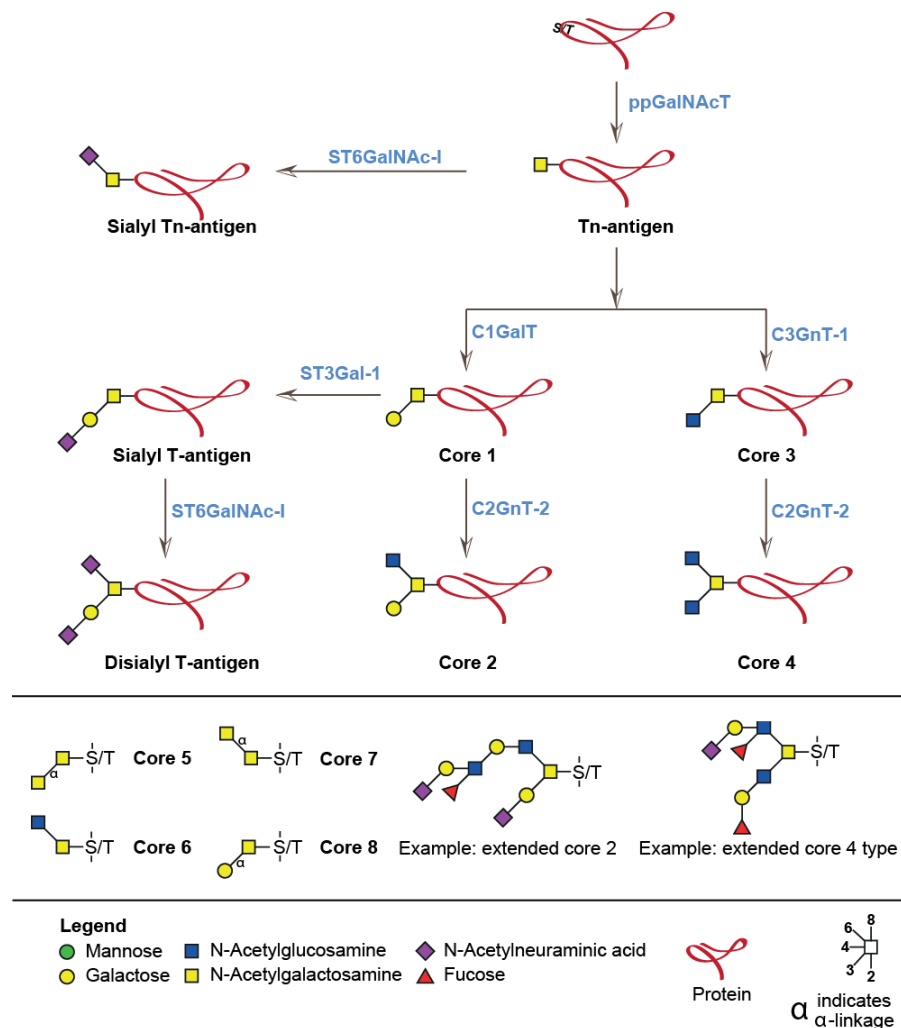


Figure 1.4: O-Glycan Biosynthesis: Biosynthesis pathway of the four most reported mucin type O-glycan core structures. *ppGalNAcT* enzymes initiate O-glycosylation by transferring a GalNAc to Ser or Thr residues. *C1GalT* adds a Gal residue to synthesise the Core 1 structure, which can be branched by *C2GnT* to form Core 2. The Core 3 structure is synthesised by *C3GnT* that adds GlcNAc to GalNAc. Core 3 can be branched by *C2GnT2* to form Core 4. These structures can be further elongated, sialylated or sulphated.

1.1.3. Biological Significance of protein glycosylation - all theories are correct!

Glycoproteins play a major role in numerous biological processes such as cell-cell communication and interaction, immune response, intracellular targeting, intracellular recognition, fertilization, inflammation, embryonic development, viral replication and parasitic infection (Figure 1.5) (25-28). Variations in glycoprotein macro- and micro-heterogeneity are key indicators of cellular physiological

conditions and as such, of disease. Glycoproteins containing multiple sites of glycosylation frequently exhibit different glyco-profiles on the individual sites (16, 29-32), and depending on the protein and disease, glycan-alterations can either just occur on selective positions or affect the global glycoprofile of a protein and/or tissue. These variations in both micro and macro-heterogeneity influence the physiological functions of the glycoproteins.

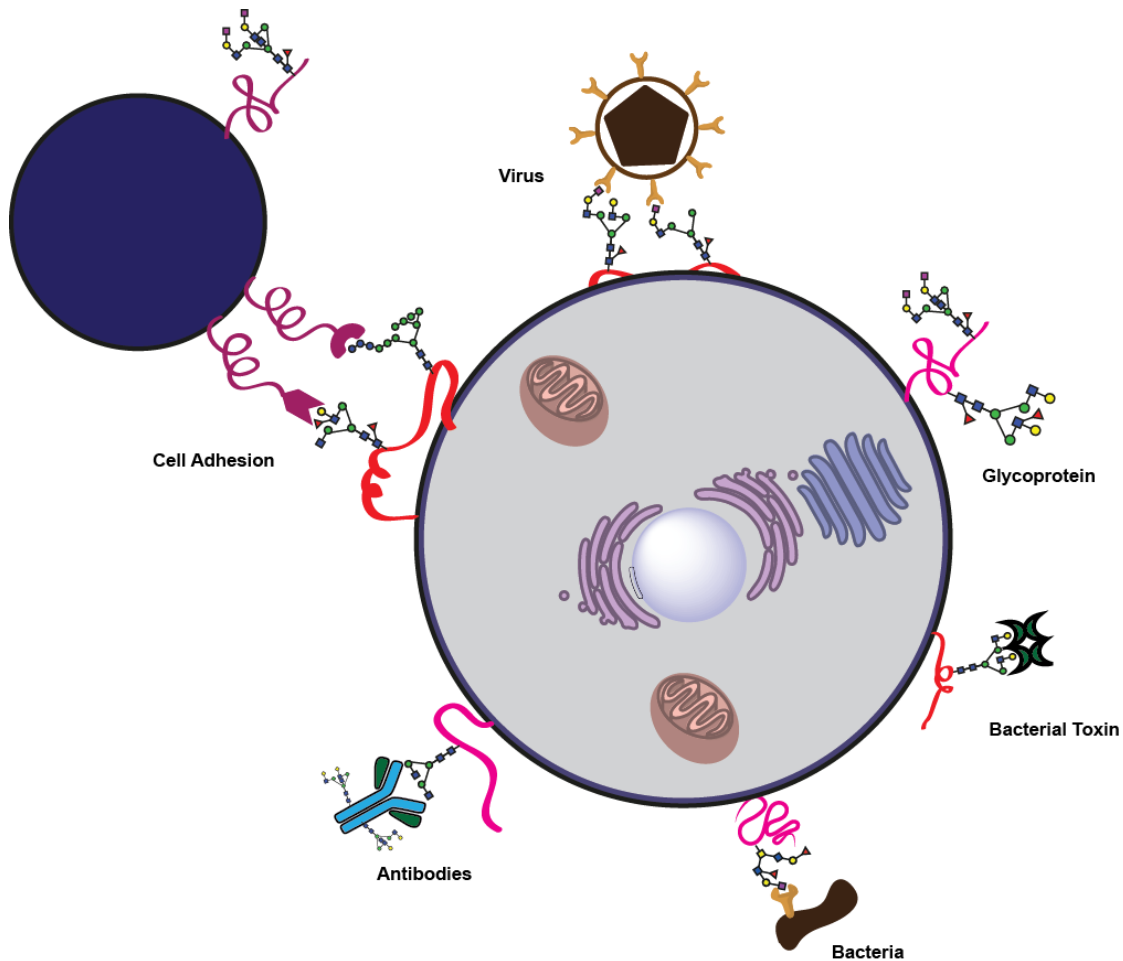


Figure 1.5: Glycans as recognition signals [illustration adapted from (61)]: schematic illustration depicting various protein–glycan interactions at the cell surface mediating cell–cell binding, cell–microbe (bacterial, viral and bacterial toxin) adhesion and cell–antibody binding.

Glycosylation plays a key role in protein folding, quality control, stability and protein targeting during protein translation (33). Inherited deficiencies in the *N*-glycosylation biosynthetic pathway thus can result in severe life threatening clinical manifestations known as congenital disorders of glycosylation (CDGs) (34). Defects in key components of the glycosylation machinery affect not only that

enzyme, but affect essentially all glycoproteins that are subjected to the secretory pathway, thus interfering with cellular interactions on a global level.

Immunoglobulins (Ig's) and their respective Fc receptors belong to one of the best studied and understood glycoproteins as they are key components of both, the innate and adaptive immune systems (35). The specific glycosylation of immunoglobulins and their receptors has been described to be crucial for maintaining and modulating effector functions. The site-specific presence of an oligomannose-type *N*-glycan on IgE is necessary for initiating anaphylaxis (36). Another striking example is the structure of the Fc fragment of immunoglobulin G (IgG) (37-41). IgG sialylation and core fucosylation levels are known to modulate anti- and proinflammatory properties of this molecule (42, 43). In addition, the composition of Fc *N*-glycan changes in certain disease state (44), affect affinity for many FcγRs (45, 46) and influences the Fc quaternary structure (Fc glycan provides the IgG-Fc region a more open conformation, allowing the binding of FcγR) (47, 48).

Glycan micro-heterogeneity can significantly affect glycoprotein properties and trigger different immunological responses as best exemplified by the ABO blood group system (49, 50). The ABH antigens are complex terminal glycan epitopes present on glycoconjugates that are generated by the addition of *N*-acetylgalactosamine (A antigen) or D-galactose (B antigen) to existing *N*-glycan, O-glycan or glycolipid structures via the action of the respective ABO glycosyltransferases. Even the orientation (linkage) of the attached oligosaccharide with respect to the polypeptide may greatly affect the properties of individual glycoproteins and trigger different immunological responses.

Over the past years it has become evident that defects in the attachment of carbohydrates to proteins are associated in a number of human diseases such as (but not limited to) cancer, inflammation, Alzheimer's disease, diabetes, cystic fibrosis or metabolic disorders (25, 26, 34, 51-60). Therefore, identification, structural characterisation and relative quantification of these glycoproteins and their individual glycosylation signatures is an essential prerequisite to understand the various complex biological process that glycoproteins are involved in.

1.2. MS BASED GLYCOPROTEIN ANALYSIS

The comprehensive characterisation of glycosylation in a protein specific manner is crucial to subsequently understand the influence glycosylation has on the biological functions of the individual proteins (functional glycoproteomics) (62). Mass Spectrometry (MS) based approaches, frequently applied in combination with nano-scale liquid chromatography (nano-LC), have emerged as indispensable tools to identify, characterise or quantify complex glycoproteins due to their universal applicability, sensitivity and high throughput capabilities. These facilitate the identification and in-depth characterisation of glycoproteins, glycosylation sites and glycan structures (63, 64). Furthermore, tandem MS provides fragmentation data of selected precursor ions that assist in identification and characterisation of the peptide/protein sequences, glycosylation sites and glycan structures even from highly complex mixtures subject to appropriate sample preparation.

Determination of the *in vivo* functions of glycoproteins is usually not a simple task. One of the major analytical challenges in MS based glycoprotein analysis is the propensity of glycans to form many different isomers from a comparably very limited set of building blocks (65). Adding to the overall complexity associated with glycan structures, the determination of glycoprotein micro- and macro-heterogeneity presents the second major analytical challenge that needs to be overcome to provide the analytical basis to study glycoprotein function. Additional challenges are introduced by the fact that disease associated changes occurring on glycoproteins are often present at low levels (66). Consequently multi-dimensional analytical strategies are a necessity for glycoprotein characterisation. The comprehensive analysis of glycoprotein involves: (i) Protein level - identification of the carrier protein (ii) Glycan level – identification of the glycans attached to the carrier protein including monosaccharide composition, sequence structure, and (iii) site level – identification of the sites of glycosylation and individual macro- and micro-heterogeneity (Table 1.1) Achieving these goals requires the development of novel sensitive, robust and precise methods that allow covering one or more of the aforementioned aspects for both glycan and glycopeptide analysis within a single experiment and from a minimum of initial sample material.

Table 1.1: Comparison of different MS based approaches addressing various aspects of glycoprotein analysis

Protein Focused Glycoproteomics	Glycan focused glycoproteomics	Glycoprotein focused glycoproteomics
<p>Pro's:</p> <ul style="list-style-type: none"> ➤ ID of previously <i>N</i>-glycosylated proteins 	<p>Pro's:</p> <ul style="list-style-type: none"> ➤ Uncovers structure details ➤ Provides global structural overview 	<p>Pro's:</p> <ul style="list-style-type: none"> ➤ Simultaneous investigation of peptide and glycan moieties ➤ Site specific structure data ➤ Assignment to individual proteins possible
<p>Con's:</p> <ul style="list-style-type: none"> ➤ Information on <i>N</i>-glycan structures is lost ➤ No information on <i>O</i>-glycosylated sites 	<p>Con's:</p> <ul style="list-style-type: none"> ➤ No site specific information ➤ No assignment to individual glycosylation site possible 	<p>Con's:</p> <ul style="list-style-type: none"> ➤ Isobaric glycan modifications hard to differentiate

1.3. GLYCOMICS

Glycomics determines qualitative and quantitative changes in the glycome after the glycans are removed from the proteins (67-69). Despite the fact that for most glycoproteins the glycome does not reflect any site specific distribution on the respective protein (unless they just contain a single site of glycosylation in their protein sequence), these analyses provide a deeper insight into the peculiarities of the cellular glycosylation machinery and the individual structures present on a glycoprotein or glycoprotein mixture (62). Typically glycan analyses involve numerous sample preparation steps starting with the release of protein bound glycans followed by subsequent isolation of the released glycans and, if necessary, derivatisation prior MS analysis. Depending upon the linkage chemistry between the oligosaccharide and peptide backbone either chemical or enzymatic methods can be applied to release intact glycans from a glycoprotein (Figure 1.6). The enzyme peptide- N^4 -(*N*-acetyl- β -glucosaminyl) asparagine amidase F (PNGase F) is universally applicable for the release of *N*-linked glycans due to its broad selectivity and ability to cleave the amide bond between the innermost core GlcNAc and asparagine residue within the peptide backbone. However, *N*-glycans carrying a so called core α 1,3 fucose modification on the innermost GlcNAc are

resistant to PNGase F and require a different enzyme, PNGase A, to be released from the peptide backbone (70).

Enzymatic cleavage opportunities are limited for the release of O-glycans, also possibly due the lack of a consensus core structure. Usage of Endo- α -N-acetylgalactosaminidase ("O-glycanase") is restricted due to its ability to cleave only the amide bond between an unmodified core-1 O-glycan and Ser/Thr peptide backbone (71). Thus chemical methods are the preferred option if a global release of all O-glycan structures is envisaged. Various chemical approaches such as reductive (72) and non-reductive β -elimination (73, 74) can be applied to release O-glycans. In the case of hydrazinolysis, both N- and O-linked glycans can be released sequentially depending upon the applied release conditions (75, 76).

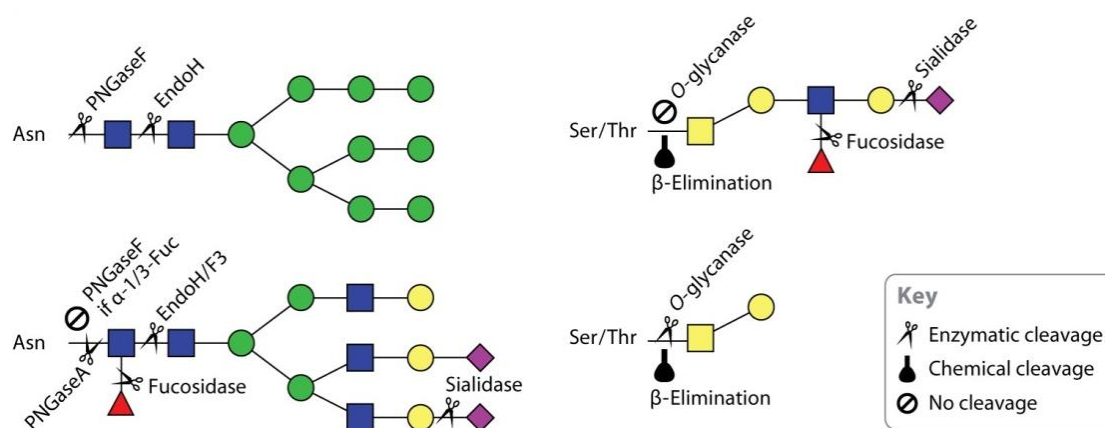


Figure 1.6: Representation of various chemical and enzymatic methods available for glycan release from glycoproteins showing the site of action. Figure adapted from (77).

1.3.1. Global glycomics

Matrix Assisted Laser Desorption/Ionisation (MALDI) and Electrospray Ionisation (ESI) are both widely used ionisation techniques for the analysis of released glycans. MALDI-MS based glycan analysis often involves permethylation of the released glycan to improve the ionisation efficiency and to stabilise the glycan substituents such as sialic acids that are comparably labile under MALDI ionisation conditions. However, glycan permethylation requires a number of additional sample preparation steps that can result in sample losses during purification and the introduction of artefacts such as incomplete permethylation if not performed properly (41).

Considering the limitations for glycan mass profiling by MALDI-MS, ESI-MS provides the additional advantage for easy implementation of online coupling with various LC-separation approaches that provide an additional dimension of information, which allows to separate isobaric glycan structures prior MS analysis that otherwise are difficult or impossible to differentiate by simple MS based techniques. LC-MS has been used in various configurations for glycan analysis including normal phase (NP), reversed-phase (RP), porous graphitized carbon (PGC), hydrophilic interaction (HILIC), ion-exchange (IEX), and high pH anion exchange (HPAEC) chromatography. Depending on the choice of analytical method used, reducing end glycan derivatisation might be necessary introducing additional sample preparation steps.

1.3.2. Tandem Mass Spectrometry provides deeper structural insights

Over the past years PGC has emerged as one of the most widely used chromatographic media for LC-separation of non-derivatised oligosaccharides due to its unique features that allow separating isobaric glycans and thus enabling detailed structural glycan characterisation within a single experiment (78-80). PGC-LC-ESI-MS/MS glycomics analyses of *N*- and *O*- glycan alditols can be performed to produce either negatively (81-83) or positively charged ions (79, 84, 85).

In general, Collision Induced Dissociation (CID) fragmentation of negative ion precursors produces prominent C-type glycosidic cleavages and A-type cross-ring glycosidic fragments rather than the B- and Y-type ions observed in the fragmentation of positive ion precursors (Figure 1.7) (86, 87). In addition, in positive mode CID fragmentation, glycans undergo monosaccharide rearrangements impeding structural interpretation and subsequent glycan structure assignment (88). The most important diagnostic ions observed upon CID of negatively charged molecules are highly informative D-ions (providing information on the 6 arm antennae substitutions) and E-ions (3 arm antennae). The presence or absence of a bisecting GlcNAc can be inferred from the D-221 ions. C-ions derived from the non-reducing end of the glycan provide information on the sequence of the constituent monosaccharide residues (Figure 1.8). In many cases, however, negative ion spectra are less ambiguous (highly informative

MSMS spectra and no observed monosaccharide rearrangement) than positive ion spectra (86, 89-92).

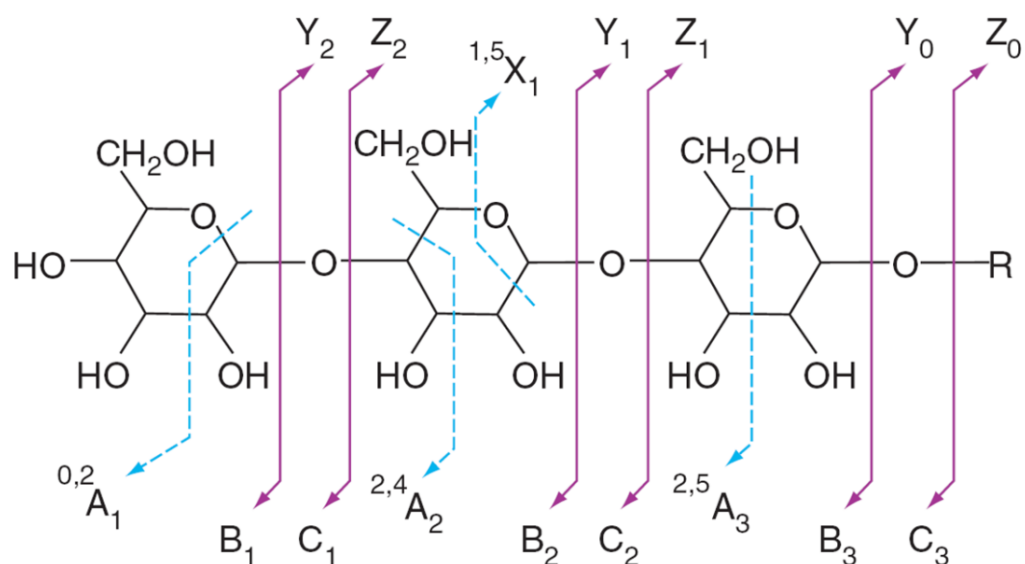


Figure 1.7: In this nomenclature (87), ions retaining the charge on the non-reducing terminus are named A, B and C, and the ions retaining charge on the reducing terminus are X, Y and Z. A and X correspond to cross-ring cleavages, whereas B, C, Y and Z correspond to glycosidic cleavages. Subscript numbers denote the cleavage position, starting at the reducing terminus for the X, Y and Z ions and at the non-reducing terminus for the others. In the case of ring cleavages, superscript numbers are given to show the cleaved bonds.

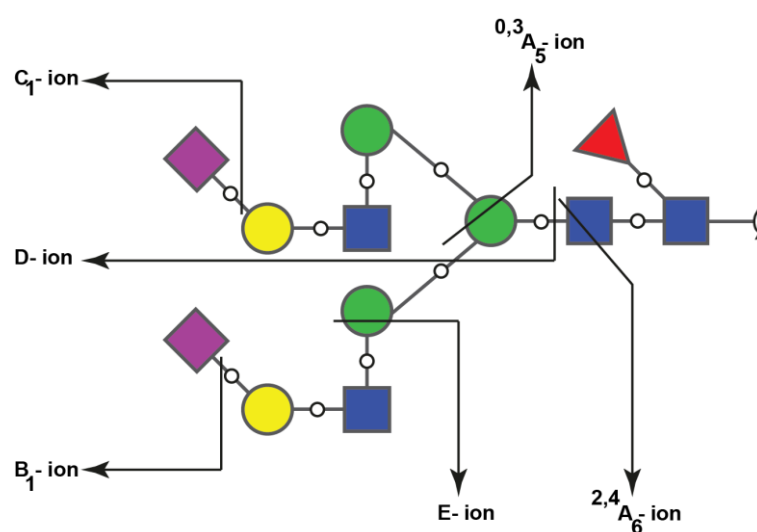


Figure 1.8: Exemplary bi-antennary N-glycan showing the major fragment ions. The oxygen atoms involved in the glycosidic bond are shown to differentiate B- and C-ions [illustration adapted from (91)].

In summary, together the information provided by these diagnostic fragment ions observed in negative mode fragmentation in combination with the separation capacity offered by the PGC enables differentiation and identification of isobaric glycan structures, which otherwise cannot be distinguished by conventional mass spectrometric techniques. Additional information supporting the structural assignment of the glycan alditol using PGC can be derived from the knowledge of the elution order of isomers that has been established using standard compounds (81-83) (Figure 1.9).

1.3.3. Glyco-Bioinformatics – the next frontier to conquer in PGC based glycomics

Despite the unique analytical opportunities provided by these technologies, interpretation of glycan fragmentation data remains to date a time consuming task posing a major limiting factor for high-throughput automated analyses due to the lack of reliable and versatile bioinformatics tools. The GlycoMod online tool available on the ExPASy Server (<http://web.expasy.org/glycomod/>) predicts potential glycan structures from experimentally determined masses (93). The GlycoWorkbench tool assists in manual interpretation and MS/MS data assignment (94). The UniCarb-DB database contains annotated MS/MS spectra of *N*- and *O*- linked glycans in addition to their retention times, and associated experimental metadata descriptions (95, 96). However, tools that allow a direct comparison of query spectra against the library spectra are still lacking in most of the mentioned software and database packages. Often critical information pertaining to MS instrumentation, experimental conditions, separation techniques, retention times, fragmentation profiles and/or structural diagnostic ion features is limited in all these glyco-bioinformatics tools, thus additional specialised tools will be required to satisfy the demand put forward by current analytics. The establishment and collection of well annotated glycan spectral libraries presents one of these opportunities that will facilitate reliable structure assignment and support data analysis automatisation.

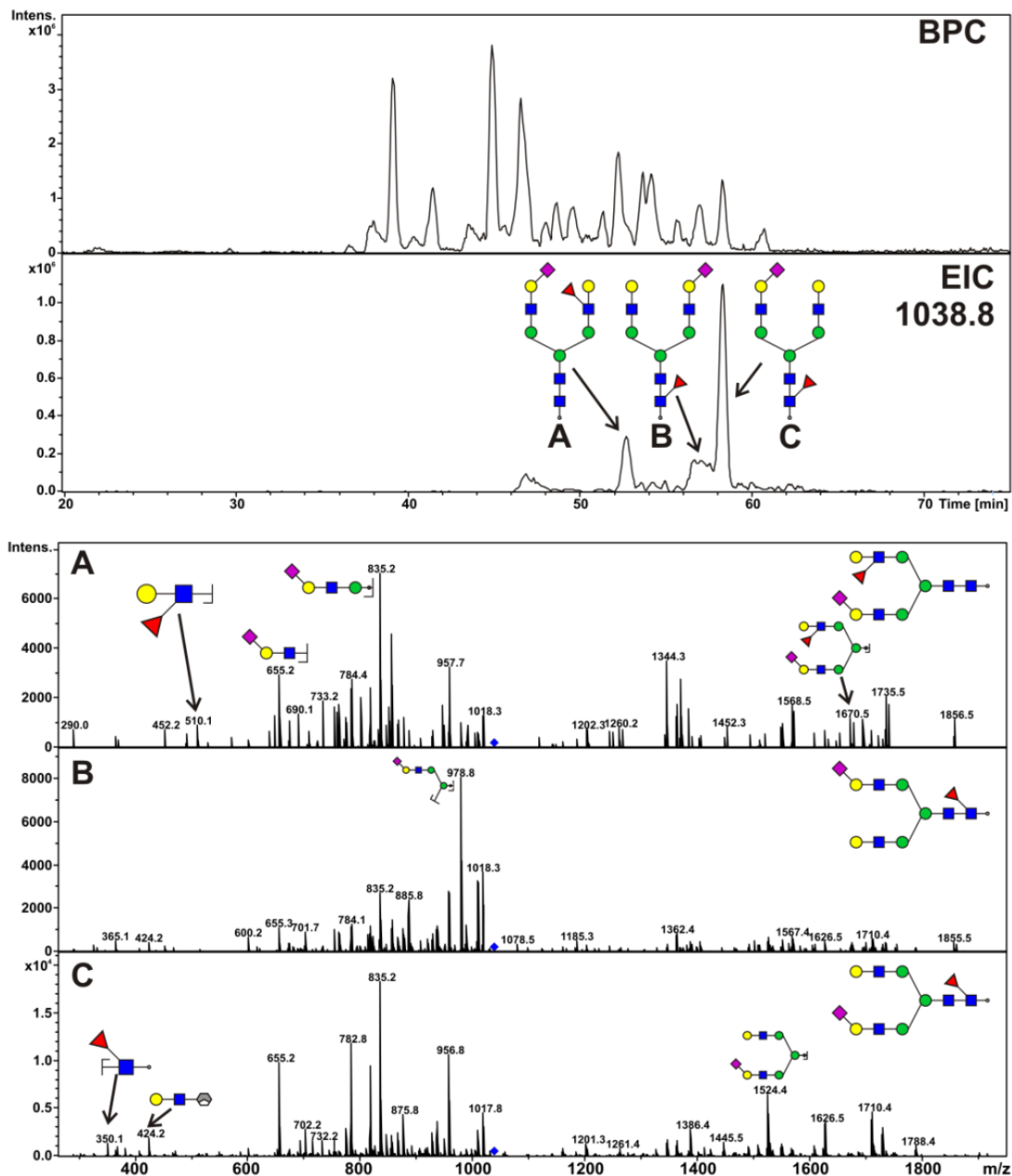


Figure 1.9: N-glycans from human secretory Immunoglobulin A (sIgA) analysed by PGC nanoLC-ESI MS/MS [figure taken from (97)]. Top panel: The base peak chromatogram (BPC) provides a global overview of the present structures. An extracted ion chromatogram (EIC), here presented for m/z 1038.8, shows three distinct isobaric N-glycan alditols with individual LC elution properties. Due to this chromatographic separation individual fingerprint MS/MS spectra can be acquired, that subsequently allow differentiation and structural characterisation (e.g. fucose linkage) of these isobaric oligosaccharide alditols.

1.3.4. Monosaccharide Linkage analysis

Oligosaccharide assembly by individual monosaccharide building blocks can occur in the same sequence but the individual building blocks can be assembled in different linkages. The type of linkage, however, significantly impacts the biological functions of these oligosaccharides and/or the carriers they are attached to, such as glycoproteins. Glycan structure assignment as well as MS based composition determination alone is frequently based on the currently available knowledge on glycan biosynthesis pathways. Nevertheless, many monosaccharide building blocks are indistinguishable by mass alone and detailed linkage information is also not easily obtained by single MS/MS analyses. Therefore, in-depth glycan analysis also requires means to determine monosaccharide composition and their respective linkages. The use of specific exoglycosidases represents one possibility to elucidate monosaccharide identity and linkage. Monitoring their activity by different analytical tools such as mass spectrometry, HPLC or Capillary Gel Electrophoresis coupled with Laser Induced Fluorescence detection (CGE-LIF) does provide information on oligosaccharide sequence, monosaccharide identity, anomericity and linkage as these enzymes are highly substrate specific (98, 99). However, this method requires a vast array of well characterised enzymes, sufficient amounts of material and last but not least it is time consuming to achieve complete glycan sequencing. In addition, this approach fails if suitable exoglycosidases are unavailable.

Numerous analytical methods have been developed to identify and quantitate monosaccharides from glycoconjugates. Several decades ago gas chromatography interfaced with electron impact ionisation mass spectrometry (GC-MS) has been established for monosaccharide determination and is still considered a state of the art method for this purpose. It is based on the conversion of monosaccharides into partially methylated alditol acetals (PMAAs) that are obtained after a series of derivatisation and hydrolysis steps: permethylation, acid hydrolysis and reduction followed by acetylation of partially methylated sugar alditols (Figure 1.10). Different types of monosaccharide residues and their linkages can be identified based upon the GC retention time and their characteristic fragmentation pattern (100-102). However, this method requires a dedicated GC-MS instrument and posing a major limitation for routine analysis in any glycoproteomics laboratory. As mentioned earlier, the biosynthetic

glycosylation machinery is often disturbed as a consequence of disease onset and progression, and this can be due to differential regulation of one or more glycosyltransferases (103). Quantitative monosaccharide and linkage analysis thus can provide novel insights into the cellular glycosylation, and subsequently contribute to a better understanding on the functional role of glycosylation.

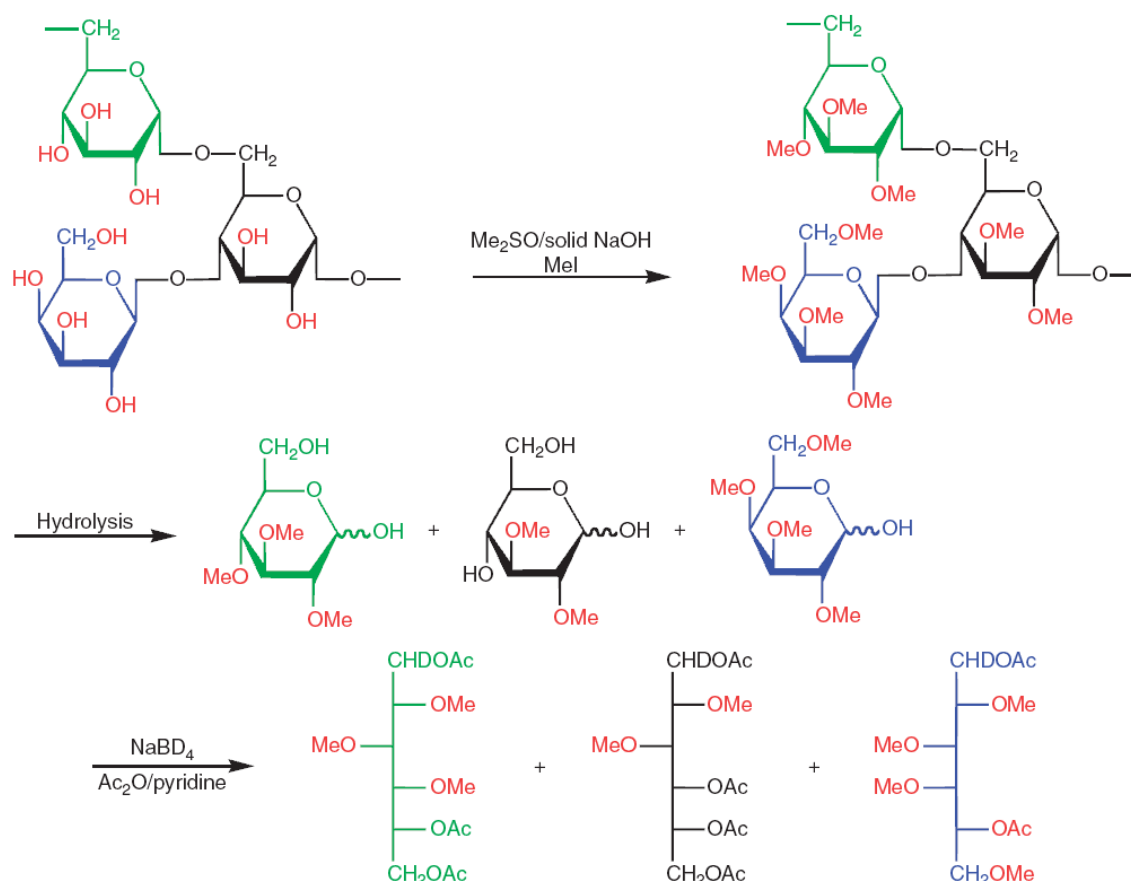


Figure 1.10: Schematic representation of a monosaccharide linkage analysis exemplified on an artificial polysaccharide

1.4. GLYCOPROTEOMICS

Glycoproteomics aims at the concomitant identification of not only the glycan composition, but also the sites of glycosylation and the identification of the protein the glycan is attached to. This information is critical for establishing the relationship between glycan structures, their individual distribution within a glycoprotein and, most importantly, their respective involvement in their protein's carrier biological function.

Mass spectrometry (MS) based glycopeptide analysis strategies rely largely on traditional “bottom-up” approaches. Typically, proteins obtained from biologically relevant samples such as serum, plasma or cell extracts have been electrophoretically separated to reduce the sample complexity, but over the past years also so called "shotgun" glycoproteomics applications have gained wider attention, in particular when combined with more or less targeted glycopeptide enrichment strategies (104-106). All of these bottom-up approaches usually involved the proteolytic digestion of the proteins. A large variety of different proteases can be used for this purpose such as trypsin, chymotrypsin, Asp-N, Glu-C or Lys-C that cut the protein into small peptides/glycopeptides and thus making them more assessable to subsequent mass spectrometric analyses (31, 107). Direct analysis of the peptide and glycopeptide mixture obtained after proteolytic digestion can be achieved by either ESI-MS/MS or MALDI-TOF-MS. In contrast to MALDI-TOF-MS, ESI-MS analysis is traditionally coupled with LC-separation devices. This combination of online LC separation (chromatographic profiling) prior mass detection provides an additional level of separation resulting in a tremendous information gain that has been shown to be superior for the analysis of complex mixtures (Figure 1.11) (31). Over the past decades, RP-LC-ESI-MS has been the most widely used technique for LC-ESI MS/MS based (glyco)proteomics (16, 108, 109).

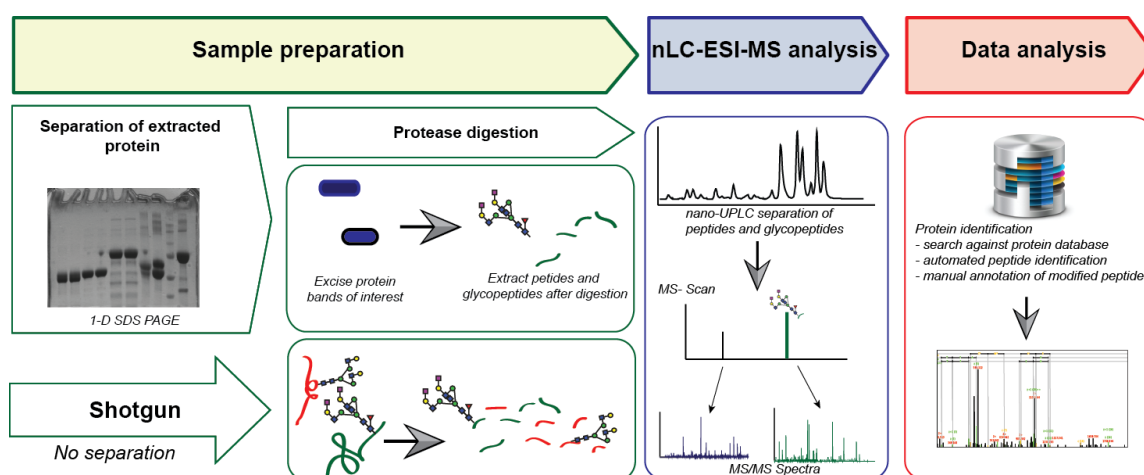


Figure 1.11: Schematic representation of a glycoproteomics workflow: Protein mixture extracted from various biologicals can either be subjected to direct proteolysis or separated electrophoretically and then subjected to proteolytic digestion. These peptide/glycopeptide fragments are analysed using LC-ESI-MS where peptides and glycopeptides can be either detected simultaneously or in separate after glycopeptide enrichment.

However, simultaneous analysis of peptides and glycopeptides can be tricky. Glycoprotein proteolysis often results in unequal mixtures of these compounds as glycopeptide microheterogeneity reduces the concentration of each individual glycopeptide molecule compared to the unmodified peptides that derive from the same digest (31). Hydrophobic molecules also tend to provide stronger signals compared to hydrophilic ones, further complicating glycopeptide detection in the presence of unmodified peptides (110). Subsequently, glycopeptide signal strengths are significantly lower compared to their unmodified counterparts, mostly due to the presence of the large hydrophilic glycan moiety (111). Therefore glycopeptide enrichment is often performed to facilitate their detection and identification (13, 112, 113).

1.4.1. Glycopeptide enrichment strategies in glycoproteomics

In contrast to the majority of glycopeptide enrichment methods Table 1.2 Hydrophilic Interaction Chromatography (HILIC) comes with the unique advantage to enable glycopeptide enrichment in a largely glycan structure unbiased manner. During the HILIC enrichment process glycopeptides are also not chemically or enzymatically altered - this is highly relevant for in-depth glycoproteomics. Another significant advantage of HILIC is that both peptide and glycan present in the enriched fraction can be analysed in a high throughput fashion as intact glycopeptides but also individually after enzymatic treatments with PNGase F/A.

Table 1.2: Comparison of various glycopeptides enrichment techniques used in a glycoproteomics workflow

	Lectins	Hydrazine	TiO ₂	HILIC	Cellulose	Carbon
Binding principle	relative affinities to glycan motifs	chemical oxidation#	charge	hydrophilicity	hydrophilicity	dispersive and charge interactions
Glycomics Compatibility	++	-	++	++	++	++
N-linked O-linked	Lectin dependent	~ both#	both¶	both*	both*¶	both
Peptide	(±)	(±)	(±)	++	++	(±)
Protein	+	-	-	(±)	(±)	-
specificity	reproducible with unknown selectivity	reproducible with unknown selectivity	selective for charged glycans	reproducible, hydrophilicity dependent	reproducible, hydrophilicity dependent	reproducible, not sugar specific

Vicinal hydroxyl groups required * Depends on size of oligosaccharide chain

¶ At present sufficient O-glycan data is not available

1.4.2. MS based Glycopeptide Characterisation

Tandem Mass Spectrometry (MS/MS) is a powerful tool due to its ability to derive structural information about glycans and peptides that also enables localisation of glycosylation sites. This is mainly achieved by recent advancements in various fragmentation techniques such as Electron transfer dissociation (ETD) and Electron-Capture Dissociation (ECD) that are now frequently used as alternatives to well established Collision Induced Dissociation (CID).

For glycopeptides, low-energy vibrational activation (CID) results in the preferential fragmentation of the carbohydrate moiety, usually with little or no peptide backbone cleavage. High-energy collision dissociation (HCD) allows cleavage of both peptide bonds as well the glycosidic ones, which provides information on peptide sequence and glycan structure in a single experiment. Under optimal collision energy settings, HCD fragmentation of glycopeptides results in distinct Y1 ions (peptide+GlcNAc) allowing effective glycopeptide identification (114, 115). ECD and ETD fragmentation, however, predominantly produce c' ions and z'

radical ions resulting from the cleavage of the N–C α bond within a peptide (Figure 1.12). During this type of fragmentation, PTMs still remain intact and attached to the amino acid, thus allowing identification of the site of modification within a peptide sequence (116, 117). Combination of the two complementary fragmentation techniques ETD/ECD with CID thus provides investigators with the opportunity to characterise both glycans and their site of attachment within a single LC-MS/MS experiment (118). However, still some technical and analytical challenges need to be overcome in ETD/ECD glycopeptide fragmentation to make this an even more effective technique, especially if larger modifications such as *N*-glycans are to be analysed by this approach.

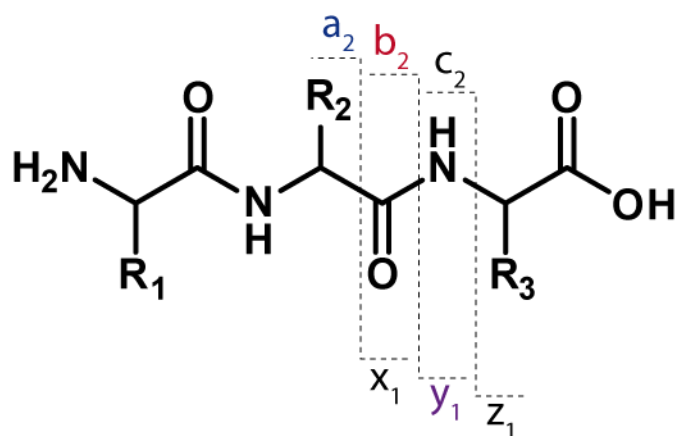


Figure 1.12: Peptide fragmentation notation using the scheme described in (119). Peptide fragment ions are indicated by *a*, *b*, or *c* if the charge is retained on the N-terminus and by *x*, *y* or *z* if the charge is maintained on the C-terminus. The subscript indicates the number of amino acid residues in the fragment.

1.5. STATE OF THE ART GLYCOPEPTIDE SYNTHESIS

The ability to produce defined and well-characterised compounds by synthetic approaches has always been a major driving force in science. The establishment of synthetic routes allowing the production of defined DNA or peptide sequences in high yields and purity has been crucial for the development of biosciences and our current understanding of cellular functions (120, 121). However, to date the synthetic capacities to easily produce a large variety and quantity of synthetic glycopeptides is still lagging behind classical peptide or DNA synthesis strategies, also because of the large structure complexity associated with glycans.

The capacity to produce glycopeptides with defined larger glycan structures exhibiting an *in vivo* glycan-peptide linkage also represents an important step forward towards the development of potential glycopeptide-based diagnostics and therapeutics and will also significantly advance the possibilities in functional glycoproteomics (122).

Synthesis of glycopeptides with specific structures requires the combinatorial approach of both carbohydrate and peptide chemistry. Over the decades, several strategies have in principle been established for the synthesis of homogenous glycopeptides/proteins (123-126). During glycopeptide synthesis, the polypeptide backbone can be synthesised using well established peptide chemistry either in solution or on a solid support to which a carbohydrate moiety can be introduced at different levels during the synthesis. The most crucial step in glycopeptide synthesis is the formation of either *N*- or *O*-glycosidic linkages between the saccharide donor and the relevant amino acid in the polypeptide backbone. In order to achieve this, two approaches are being used: the direct glycosylation of a properly protected full-length peptide (convergent synthesis) or the use of a preformed glycosylated amino acid building block for glycopeptide synthesis (Stepwise synthesis) (**Figure** Figure 113). In the stepwise glycopeptide synthesis approach, the carbohydrate moiety is first attached to either an 9-fluorenylmethyloxycarbonyl(Fmoc)- or *tert*-butoxycarbonyl(*t*-boc)-protected Asn residue. Later, this glycosylated amino acid (GAA) is then used in solid phase glycopeptide synthesis (SGPSS) (127). However, glycopeptide synthesis has not become as routine, yet, because of the difficulty obtaining the necessary various glycosylated amino acid units (128).

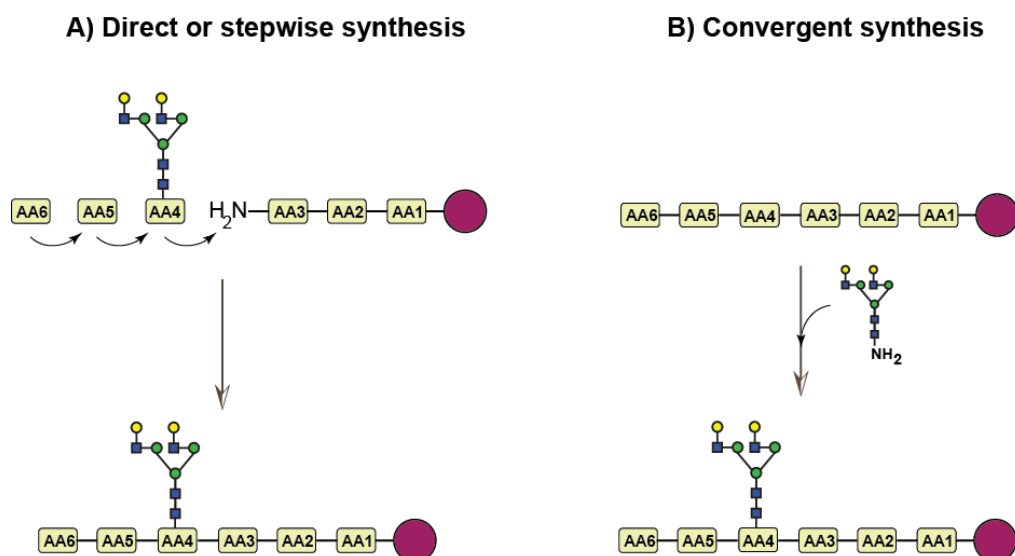


Figure 1.13: Two different approaches used for glycopeptide synthesis (a) Direct or stepwise synthesis and (b) Convergent synthesis.

Despite tremendous advances in carbohydrate chemistry (129-131) the production of large oligosaccharides with diverse building blocks still requires substantial time and material resources, in particular if an *in vivo* like linkage between an oligosaccharide and an amino acid are required. In addition, the introduction of certain biologically important glyco-features such as α -fucose, α -sialyl or β -mannose linkages between the monosaccharide building blocks are challenging and represent major limiting steps in the production of sufficient quantities of larger oligosaccharides such as *N*-glycans and make this route with the current technologies unfeasible. Despite these challenges, total complete synthesis of glycopeptide bearing fucosylated biantennary, disialylated *N*-glycans was in principle accomplished by the Danishefsky research group (132), though with considerable efforts that yet make it unfeasible for routine production.

One way of circumventing these obstacles is to make use of nature's glycosylation potential by isolating the compounds of interest from natural resources such as egg yolk (133-137). These compounds can easily be purified and subsequently transformed into Fmoc- or *t*-boc protected building blocks, which then can be applied in well-established solid phase peptide synthesis (133-137) (Figure 1.14). This approach is significantly more cost effective, quicker and allows the production of g-quantities within just 2-3 weeks.

The Fmoc strategy has been widely used for glycopeptide synthesis over Boc strategy because of milder deprotection conditions. Thus having a methodology in hand that allows easy production of the necessary glycosylated amino acid building blocks to selectively produce defined glycopeptides will significantly contribute towards better understanding of the biological role of glycoconjugates and their analysis.

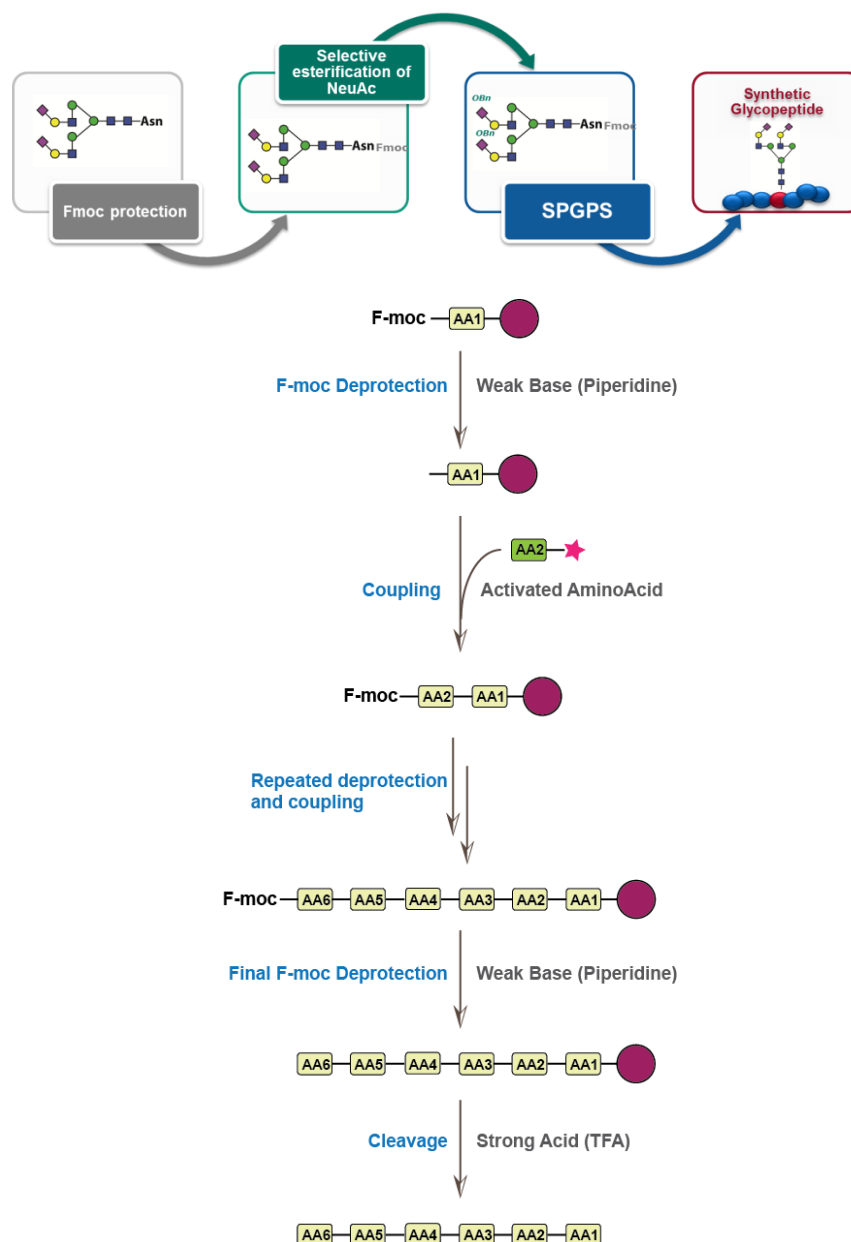


Figure 1.14: Schematic representation of major steps involved in solid phase glycopeptide synthesis (SPGPS) using Fmoc chemistry.

1.6. OBJECTIVE OF MY THESIS

The main objective of my PhD thesis was to develop novel analytical methods and technologies for high-throughput automated glycomics and glycoproteomics. These objectives were achieved using an interdisciplinary approach that included aspects of synthetic peptide chemistry, analytical biochemistry and in particular mass spectrometry and liquid chromatography based technologies that I combined to tackle the challenges listed below:

- 1) To develop fully customisable, easily adaptable spectral matching tool for automated glycan analysis based on PGC-LC-ESI-MS/MS.
- 2) To develop a novel, ultra-sensitive method approach for quantitative carbohydrate composition and linkage analysis using LC-ESI-MS/MS.
- 3) Develop a novel and environmentally friendly workflow for the isolation of glycosylated amino acids from natural resources and transform them into Fmoc-protected glycosylated Asn-amino acids for solid phase glycopeptide synthesis.
- 4) Synthesise a defined glycopeptide library with customisable variations in the glycan and peptide moieties of a glycopeptide.
- 5) Apply the established glycopeptide library for the systematic evaluation of sample preparation and glycopeptide analysis parameters using MS to improve the sensitivity, selectivity and accuracy in glycoproteomics research so that they can subsequently be applied onto biologically relevant samples in glyco-biomarker research.

1.7. REFERENCES

1. Varki A, Sharon N. Historical Background and Overview. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al., editors. *Essentials of Glycobiology*. 2nd ed. Cold Spring Harbor (NY)2009.
2. Stanley P. Golgi glycosylation. *Cold Spring Harb Perspect Biol*. 2011;3(4).
3. Hart GW, West CM. Nucleocytoplasmic Glycosylation. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al., editors. *Essentials of Glycobiology*. 2nd ed. Cold Spring Harbor (NY)2009.
4. Moremen KW, Tiemeyer M, Nairn AV. Vertebrate protein glycosylation: diversity, synthesis and function. *Nat Rev Mol Cell Biol*. 2012;13(7):448-62.
5. Turnbull JE, Field RA. Emerging glycomics technologies. *Nat Chem Biol*. 2007;3(2):74-7.
6. Nelson NJ. Draft human genome sequence yields several surprises. *J Natl Cancer I*. 2001;93(7):493-.
7. Raman R, Raguram S, Venkataraman G, Paulson JC, Sasisekharan R. Glycomics: an integrated systems approach to structure-function relationships of glycans. *Nature methods*. 2005;2(11):817-24.
8. Apweiler R, Hermjakob H, Sharon N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim Biophys Acta*. 1999;1473(1):4-8.
9. Dell A, Galadari A, Sastre F, Hitchen P. Similarities and differences in the glycosylation mechanisms in prokaryotes and eukaryotes. *Int J Microbiol*. 2010;2010:148178.
10. Spiro RG. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology*. 2002;12(4):43R-56R.
11. Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al. *Essentials of Glycobiology*, 2nd edition. 2 ed. New York: Cold Spring Harbor Laboratory Press; 2009. 784 p.
12. Dell A, Morris HR. Glycoprotein structure determination by mass spectrometry. *Science*. 2001;291(5512):2351-6.
13. Budnik BA, Lee RS, Steen JA. Global methods for protein glycosylation analysis by mass spectrometry. *Biochimica et biophysica acta*. 2006;1764(12):1870-80.
14. van Kooyk Y, Rabinovich GA. Protein-glycan interactions in the control of innate and adaptive immune responses. *Nature immunology*. 2008;9(6):593-601.
15. Ruddock LW, Molinari M. N-glycan processing in ER quality control. *Journal of cell science*. 2006;119(Pt 21):4373-80.
16. Kolarich D, Weber A, Turecek PL, Schwarz HP, Altmann F. Comprehensive glyco-proteomic analysis of human alpha1-antitrypsin and its charge isoforms. *Proteomics*. 2006;6(11):3369-80.
17. Stanley P, Schachter H, Taniguchi N. N-Glycans. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al., editors. *Essentials of Glycobiology*. 2nd ed. Cold Spring Harbor (NY)2009.
18. Ellgaard L, Helenius A. Quality control in the endoplasmic reticulum. *Nat Rev Mol Cell Biol*. 2003;4(3):181-91.
19. Lamriben L, Graham JB, Adams BM, Hebert DN. N-Glycan-based ER Molecular Chaperone and Protein Quality Control System: The Calnexin Binding Cycle. *Traffic*. 2016;17(4):308-26.

20. Kornfeld R, Kornfeld S. Assembly of asparagine-linked oligosaccharides. *Annu Rev Biochem.* 1985;54:631-64.
21. Hanisch FG. O-glycosylation of the mucin type. *Biol Chem.* 2001;382(2):143-9.
22. Moremen KW, Tiemeyer M, Nairn AV. Vertebrate protein glycosylation: diversity, synthesis and function. *Nat Rev Mol Cell Biol.* 2012;13(7):448-62.
23. Rivinoja A, Kokkonen N, Kellokumpu I, Kellokumpu S. Elevated Golgi pH in breast and colorectal cancer cells correlates with the expression of oncofetal carbohydrate T-antigen. *Journal of Cellular Physiology.* 2006;208(1):167-74.
24. Rivinoja A, Hassinen A, Kokkonen N, Kauppila A, Kellokumpu S. Elevated Golgi pH Impairs Terminal N-Glycosylation by Inducing Mislocalization of Golgi Glycosyltransferases. *Journal of Cellular Physiology.* 2009;220(1):144-54.
25. Dwek RA. Glycobiology: "towards understanding the function of sugars". *Biochemical Society transactions.* 1995;23(1):1-25.
26. Varki A. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology.* 1993;3(2):97-130.
27. Varki A. Biological Roles of Oligosaccharides - All of the Theories Are Correct. *Glycobiology.* 1993;3(2):97-130.
28. Varki A. Biological Roles of Glycans: Two Decades Later. *Glycobiology.* 2014;24(11):1086-7.
29. Deshpande N, Jensen PH, Packer NH, Kolarich D. GlycoSpectrumScan: fishing glycopeptides from MS spectra of protease digests of human colostrum sIgA. *J Proteome Res.* 2010;9(2):1063-75.
30. Kolarich D, Altmann F, Sunderasan E. Structural analysis of the glycoprotein allergen Hev b 4 from natural rubber latex by mass spectrometry. *Biochim Biophys Acta.* 2006;1760(4):715-20.
31. Kolarich D, Jensen PH, Altmann F, Packer NH. Determination of site-specific glycan heterogeneity on glycoproteins. *Nat Protoc.* 2012;7(7):1285-98.
32. Kolarich D, Weber A, Pabst M, Stadlmann J, Teschner W, Ehrlich H, et al. Glycoproteomic characterization of butyrylcholinesterase from human plasma. *Proteomics.* 2008;8(2):254-63.
33. Helenius A, Aebi M. Roles of N-linked glycans in the endoplasmic reticulum. *Annu Rev Biochem.* 2004;73:1019-49.
34. Ohtsubo K, Marth JD. Glycosylation in cellular mechanisms of health and disease. *Cell.* 2006;126(5):855-67.
35. Hayes JM, Cosgrave EF, Struwe WB, Wormald M, Davey GP, Jefferis R, et al. Glycosylation and Fc receptors. *Curr Top Microbiol Immunol.* 2014;382:165-99.
36. Shade KTC, Platzer B, Washburn N, Mani V, Bartsch YC, Conroy M, et al. A single glycan on IgE is indispensable for initiation of anaphylaxis. *Journal of Experimental Medicine.* 2015;212(4):457-67.
37. Kaneko Y, Nimmerjahn F, Ravetch EV. Anti-inflammatory activity of immunoglobulin G resulting from Fc sialylation. *Science.* 2006;313(5787):670-3.
38. Dwek RA, Lellouch AC, Wormald MR. Glycobiology - the Function of Sugar in the Igg Molecule. *Journal of Anatomy.* 1995;187:279-92.
39. Axford J. Glycobiology and medicine: an introduction. *Journal of the Royal Society of Medicine.* 1997;90(5):260-4.
40. Anthony RM, Ravetch JV. A novel role for the IgG Fc glycan: the anti-inflammatory activity of sialylated IgG Fcs. *Journal of clinical immunology.* 2010;30 Suppl 1:S9-14.
41. Kolarich D, Lepenies B, Seeberger PH. Glycomics, glycoproteomics and the immune system. *Curr Opin Chem Biol.* 2012;16(1-2):214-20.

42. Kaneko Y, Nimmerjahn F, Ravetch JV. Anti-inflammatory activity of immunoglobulin G resulting from Fc sialylation. *Science*. 2006;313(5787):670-3.
43. Masuda K, Kubota T, Kaneko E, Iida S, Wakitani M, Kobayashi-Natsume Y, et al. Enhanced binding affinity for FcγRIIIa of fucose-negative antibody is sufficient to induce maximal antibody-dependent cellular cytotoxicity. *Mol Immunol*. 2007;44(12):3122-31.
44. Parekh RB, Dwek RA, Sutton BJ, Fernandes DL, Leung A, Stanworth D, et al. Association of rheumatoid arthritis and primary osteoarthritis with changes in the glycosylation pattern of total serum IgG. *Nature*. 1985;316(6027):452-7.
45. Shields RL, Lai J, Keck R, O'Connell LY, Hong K, Meng YG, et al. Lack of fucose on human IgG1 N-linked oligosaccharide improves binding to human FcγRIII and antibody-dependent cellular toxicity. *J Biol Chem*. 2002;277(30):26733-40.
46. Lux A, Yu XJ, Scanlan CN, Nimmerjahn F. Impact of Immune Complex Size and Glycosylation on IgG Binding to Human FcγRs. *Journal of Immunology*. 2013;190(8):4315-23.
47. Krapp S, Mimura Y, Jefferis R, Huber R, Sondermann P. Structural analysis of human IgG-Fc glycoforms reveals a correlation between glycosylation and structural integrity. *J Mol Biol*. 2003;325(5):979-89.
48. Subedi GP, Hanson QM, Barb AW. Restricted motion of the conserved immunoglobulin G1 N-glycan is essential for efficient FcγRIIIa binding. *Structure*. 2014;22(10):1478-88.
49. Landsteiner K, van der Scheer J. On the Antigens of Red Blood Corpuscles : The Question of Lipoid Antigens. *The Journal of experimental medicine*. 1925;41(3):427-37.
50. Alfaro JA, Zheng RB, Persson M, Letts JA, Polakowski R, Bai Y, et al. ABO(H) blood group A and B glycosyltransferases recognize substrate via specific conformational changes. *J Biol Chem*. 2008;283(15):10097-108.
51. Pang PC, Chiu PCN, Lee CL, Chang LY, Panico M, Morris HR, et al. Human Sperm Binding Is Mediated by the Sialyl-Lewis(x) Oligosaccharide on the Zona Pellucida. *Science*. 2011;333(6050):1761-4.
52. Wassarman PM. The Sperm's Sweet Tooth. *Science*. 2011;333(6050):1708-9.
53. Haltiwanger RS. Regulation of signal transduction pathways in development by glycosylation. *Curr Opin Struct Biol*. 2002;12(5):593-8.
54. Muramatsu T. Carbohydrate signals in metastasis and prognosis of human carcinomas. *Glycobiology*. 1993;3(4):291-6.
55. Fukuda M. Possible roles of tumor-associated carbohydrate antigens. *Cancer Res*. 1996;56(10):2237-44.
56. Furukawa K, Takamiya K, Okada M, Inoue M, Fukumoto S. Novel functions of complex carbohydrates elucidated by the mutant mice of glycosyltransferase genes. *Biochim Biophys Acta*. 2001;1525(1-2):1-12.
57. Fuster MM, Esko JD. The sweet and sour of cancer: Glycans as novel therapeutic targets. *Nat Rev Cancer*. 2005;5(7):526-42.
58. Varki A, Lowe JB. Biological Roles of Glycans. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al., editors. *Essentials of Glycobiology*. 2nd ed. Cold Spring Harbor (NY)2009.
59. Saeland E, van Kooyk Y. Highly glycosylated tumour antigens: interactions with the immune system. *Biochemical Society transactions*. 2011;39(1):388-92.
60. Rudd PM, Elliott T, Cresswell P, Wilson IA, Dwek RA. Glycosylation and the immune system. *Science*. 2001;291(5512):2370-6.

61. Holgersson J, Gustafsson A, Breimer ME. Characteristics of protein-carbohydrate interactions as a basis for developing novel carbohydrate-based antirejection therapies. *Immunol Cell Biol.* 2005;83(6):694-708.
62. Kolarich D, Lepenies B, Seeberger PH. Glycomics, glycoproteomics and the immune system. *Curr Opin Chem Biol.* 2012;16(1-2):214-20.
63. Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature.* 2003;422(6928):198-207.
64. Mann M, Jensen ON. Proteomic analysis of post-translational modifications. *Nature biotechnology.* 2003;21(3):255-61.
65. Laine RA. A calculation of all possible oligosaccharide isomers both branched and linear yields 1.05×10^{12} structures for a reducing hexasaccharide: the Isomer Barrier to development of single-method saccharide sequencing or synthesis systems. *Glycobiology.* 1994;4(6):759-67.
66. Marino K, Bones J, Kattla JJ, Rudd PM. A systematic approach to protein glycosylation analysis: a path through the maze. *Nat Chem Biol.* 2010;6(10):713-23.
67. Everest-Dass AV, Jin D, Thaysen-Andersen M, Nevalainen H, Kolarich D, Packer NH. Comparative structural analysis of the glycosylation of salivary and buccal cell proteins: innate protection against infection by *Candida albicans*. *Glycobiology.* 2012;22(11):1465-79.
68. de Leoz ML, An HJ, Kronewitter S, Kim J, Beecroft S, Vinall R, et al. Glycomic approach for potential biomarkers on prostate cancer: profiling of N-linked glycans in human sera and pRNS cell lines. *Disease markers.* 2008;25(4-5):243-58.
69. Mechref Y, Hu Y, Garcia A, Hussein A. Identifying cancer biomarkers by mass spectrometry-based glycomics. *Electrophoresis.* 2012;33(12):1755-67.
70. Tretter V, Altmann F, Marz L. Peptide-N4-(N-acetyl-beta-glucosaminyl)asparagine amidase F cannot release glycans with fucose attached alpha 1----3 to the asparagine-linked N-acetylglucosamine residue. *European journal of biochemistry / FEBS.* 1991;199(3):647-52.
71. Bhavanandan VP, Umemoto J, Davidson EA. Characterization of an endo-alpha-N-acetyl galactosaminidase from *Diplococcus pneumoniae*. *Biochemical and biophysical research communications.* 1976;70(3):738-45.
72. Carlson DM. Oligosaccharides isolated from pig submaxillary mucin. *The Journal of biological chemistry.* 1966;241(12):2984-6.
73. Karlsson NG, Packer NH. Analysis of O-linked reducing oligosaccharides released by an in-line flow system. *Analytical biochemistry.* 2002;305(2):173-85.
74. Huang Y, Mechref Y, Novotny MV. Microscale nonreductive release of O-linked glycans for subsequent analysis through MALDI mass spectrometry and capillary electrophoresis. *Analytical chemistry.* 2001;73(24):6063-9.
75. Merry AH, Neville DC, Royle L, Matthews B, Harvey DJ, Dwek RA, et al. Recovery of intact 2-aminobenzamide-labeled O-glycans released from glycoproteins by hydrazinolysis. *Analytical biochemistry.* 2002;304(1):91-9.
76. Patel T, Bruce J, Merry A, Bigge C, Wormald M, Jaques A, et al. Use of hydrazine to release in intact and unreduced form both N- and O-linked oligosaccharides from glycoproteins. *Biochemistry.* 1993;32(2):679-93.
77. Rakus JF, Mahal LK. New Technologies for Glycomic Analysis: Toward a Systematic Understanding of the Glycome. *Annual Review of Analytical Chemistry, Vol 4.* 2011;4:367-92.

78. Pabst M, Bondili JS, Stadlmann J, Mach L, Altmann F. Mass + retention time = structure: a strategy for the analysis of N-glycans by carbon LC-ESI-MS and its application to fibrin N-glycans. *Analytical chemistry*. 2007;79(13):5051-7.
79. Stadlmann J, Pabst M, Kolarich D, Kunert R, Altmann F. Analysis of immunoglobulin glycosylation by LC-ESI-MS of glycopeptides and oligosaccharides. *Proteomics*. 2008;8(14):2858-71.
80. Fan JQ, Kondo A, Kato I, Lee YC. High-performance liquid chromatography of glycopeptides and oligosaccharides on graphitized carbon columns. *Analytical biochemistry*. 1994;219(2):224-9.
81. Schulz BL, Packer NH, Karlsson NG. Small-scale analysis of O-linked oligosaccharides from glycoproteins and mucins separated by gel electrophoresis. *Anal Chem*. 2002;74(23):6088-97.
82. Thomsson KA, Karlsson NG, Hansson GC. Liquid chromatography-electrospray mass spectrometry as a tool for the analysis of sulfated oligosaccharides from mucin glycoproteins. *Journal of chromatography A*. 1999;854(1-2):131-9.
83. Karlsson NG, Wilson NL, Wirth HJ, Dawes P, Joshi H, Packer NH. Negative ion graphitised carbon nano-liquid chromatography/mass spectrometry increases sensitivity for glycoprotein oligosaccharide analysis. *Rapid communications in mass spectrometry : RCM*. 2004;18(19):2282-92.
84. Pabst M, Bondili JS, Stadlmann J, Mach L, Altmann F. Mass plus retention time = structure: A strategy for the analysis of N-glycans by carbon LC-ESI-MS and its application to fibrin N-glycans. *Analytical Chemistry*. 2007;79(13):5051-7.
85. Pabst M, Altmann F. Influence of electrosorption, solvent, temperature, and ion polarity on the performance of LC-ESI-MS using graphitic carbon for acidic oligosaccharides. *Analytical Chemistry*. 2008;80(19):7534-42.
86. Harvey DJ. Fragmentation of negative ions from carbohydrates: part 1. Use of nitrate and other anionic adducts for the production of negative ion electrospray spectra from N-linked carbohydrates. *J Am Soc Mass Spectrom*. 2005;16(5):622-30.
87. Domon B, Costello CE. A Systematic Nomenclature for Carbohydrate Fragmentations in Fab-MS Ms Spectra of Glycoconjugates. *Glycoconjugate J*. 1988;5(4):397-409.
88. Wuhler M, Deelder AM, van der Burgt YE. Mass spectrometric glycan rearrangements. *Mass Spectrom Rev*. 2011;30(4):664-80.
89. Everest-Dass AV, Abrahams JL, Kolarich D, Packer NH, Campbell MP. Structural feature ions for distinguishing N- and O-linked glycan isomers by LC-ESI-IT MS/MS. *J Am Soc Mass Spectrom*. 2013;24(6):895-906.
90. Karlsson NG, Schulz BL, Packer NH. Structural determination of neutral O-linked oligosaccharide alditols by negative ion LC-electrospray-MSn. *J Am Soc Mass Spectrom*. 2004;15(5):659-72.
91. Harvey DJ, Royle L, Radcliffe CM, Rudd PM, Dwek RA. Structural and quantitative analysis of N-linked glycans by matrix-assisted laser desorption ionization and negative ion nanospray mass spectrometry. *Anal Biochem*. 2008;376(1):44-60.
92. Harvey DJ. Fragmentation of negative ions from carbohydrates: part 2. Fragmentation of high-mannose N-linked glycans. *J Am Soc Mass Spectrom*. 2005;16(5):631-46.
93. Cooper CA, Gasteiger E, Packer NH. GlycoMod - A software tool for determining glycosylation compositions from mass spectrometric data. *Proteomics*. 2001;1(2):340-9.

94. Ceroni A, Maass K, Geyer H, Geyer R, Dell A, Haslam SM. GlycoWorkbench: A tool for the computer-assisted annotation of mass spectra of Glycans. *J Proteome Res.* 2008;7(4):1650-9.
95. Campbell MP, Nguyen-Khuong T, Hayes CA, Flowers SA, Alagesan K, Kolarich D, et al. Validation of the curation pipeline of UniCarb-DB: Building a global glycan reference MS/MS repository. *Bba-Proteins Proteom.* 2014;1844(1):108-16.
96. Hayes CA, Karlsson NG, Struwe WB, Lisacek F, Rudd PM, Packer NH, et al. UniCarb-DB: a database resource for glycomic discovery. *Bioinformatics.* 2011;27(9):1343-4.
97. Kolarich D, Packer NH. Mass spectrometry for glycomics analysis of N- and O-linked glycoproteins.
98. Maley F, Trimble RB, Tarentino AL, Plummer TH. Characterization of Glycoproteins and Their Associated Oligosaccharides through the Use of Endoglycosidases. *Analytical Biochemistry.* 1989;180(2):195-204.
99. Guttman A. Multistructure sequencing of N-linked fetuin glycans by capillary gel electrophoresis and enzyme matrix digestion. *Electrophoresis.* 1997;18(7):1136-41.
100. Björndal H, Lindberg B, Svensson S. Mass spectrometry of partially methylated alditol acetates. *Carbohydr Res.* 1967;5(4):433-40.
101. Kochetkov NK, Chizhov OS, Zolotare.Bm, Wulfson NS. Mass Spectrometry of Carbohydrate Derivatives. *Tetrahedron.* 1963;19(12):2209-&.
102. Fournet B, Strecker G, Leroy Y, Montreuil J. Gas-Liquid-Chromatography and Mass-Spectrometry of Methylated and Acetylated Methyl Glycosides - Application to the Structural-Analysis of Glycoprotein Glycans. *Anal Biochem.* 1981;116(2):489-502.
103. Drake PM, Cho W, Li B, Prakobphol A, Johansen E, Anderson NL, et al. Sweetening the pot: adding glycosylation to the biomarker discovery equation. *Clin Chem.* 2010;56(2):223-36.
104. Gulbrandsen A, Barsnes H, Kroksveen AC, Berven FS, Vaudel M. A Simple Workflow for Large Scale Shotgun Glycoproteomics. *Proteomics in Systems Biology: Methods and Protocols.* 2016;1394:275-86.
105. Domon B, Aebersold R. Options and considerations when selecting a quantitative proteomics strategy. *Nature Biotechnology.* 2010;28(7):710-21.
106. Sun BY, Ranish JA, Utleg AG, White JT, Yan XW, Lin BY, et al. Shotgun glycopeptide capture approach coupled with mass Spectrometry for comprehensive glycoproteomics. *Mol Cell Proteomics.* 2007;6(1):141-9.
107. Geyer H, Geyer R. Strategies for analysis of glycoprotein glycosylation. *Biochim Biophys Acta.* 2006;1764(12):1853-69.
108. Herndl A, Marzban G, Kolarich D, Hahn R, Boscia D, Hemmer W, et al. Mapping of *Malus domestica* allergens by 2-D electrophoresis and IgE-reactivity. *Electrophoresis.* 2007;28(3):437-48.
109. Stadlmann J, Pabst M, Kolarich D, Kunert R, Altmann F. Analysis of immunoglobulin glycosylation by LC-ESI-MS of glycopeptides and oligosaccharides. *Proteomics.* 2008;8(14):2858-71.
110. Tang L, Kebarle P. Dependence of Ion Intensity in Electrospray Mass-Spectrometry on the Concentration of the Analytes in the Electrosprayed Solution. *Anal Chem.* 1993;65(24):3654-68.
111. Stavenhagen K, Hinneburg H, Thaysen-Andersen M, Hartmann L, Varon Silva D, Fuchser J, et al. Quantitative mapping of glycoprotein micro-heterogeneity

and macro-heterogeneity: an evaluation of mass spectrometry signal strengths using synthetic peptides and glycopeptides. *J Mass Spectrom.* 2013;48(6):627-39.

112. Pasing Y, Sickmann A, Lewandrowski U. N-glycoproteomics: mass spectrometry-based glycosylation site annotation. *Biological Chemistry.* 2012;393(4):249-58.

113. An HJ, Froehlich JW, Lebrilla CB. Determination of glycosylation sites and site-specific heterogeneity in glycoproteins. *Current opinion in chemical biology.* 2009;13(4):421-6.

114. Segu ZM, Mechref Y. Characterizing protein glycosylation sites through higher-energy C-trap dissociation. *Rapid Commun Mass Sp.* 2010;24(9):1217-25.

115. Palmisano G, Larsen MR, Packer NH, Thaysen-Andersen M. Structural analysis of glycoprotein sialylation -part II: LC-MS based detection. *Rsc Adv.* 2013;3(45):22706-26.

116. Syrstad EA, Turecek F. Toward a general mechanism of electron capture dissociation. *J Am Soc Mass Spectrom.* 2005;16(2):208-24.

117. Wuhrer M, Catalina MI, Deelder AM, Hokke CH. Glycoproteomics based on tandem mass spectrometry of glycopeptides. *Journal of chromatography B, Analytical technologies in the biomedical and life sciences.* 2007;849(1-2):115-28.

118. Alley WR, Jr., Mechref Y, Novotny MV. Characterization of glycopeptides by combining collision-induced dissociation and electron-transfer dissociation mass spectrometry data. *Rapid Commun Mass Spectrom.* 2009;23(1):161-70.

119. Roepstorff P, Fohlman J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomedical mass spectrometry.* 1984;11(11):601.

120. Froehler BC, Ng PG, Matteucci MD. Synthesis of DNA via deoxynucleoside H-phosphonate intermediates. *Nucleic Acids Res.* 1986;14(13):5399-407.

121. Merrifield RB. Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *Journal of the American Chemical Society.* 1963;85(14):2149-54.

122. Kolarich D, Lepenies B, Seeberger PH. Glycomics, glycoproteomics and the immune system. *Current opinion in chemical biology.* 2012.

123. Brik A, Ficht S, Wong CH. Strategies for the preparation of homogenous glycoproteins. *Curr Opin Chem Biol.* 2006;10(6):638-44.

124. Izumi M, Dedola S, Ito Y, Kajihara Y. Chemical Synthesis of Homogeneous Glycoproteins for the Study of Glycoprotein Quality Control System. *Isr J Chem.* 2015;55(3-4):306-14.

125. Wang LX, Amin MN. Chemical and Chemoenzymatic Synthesis of Glycoproteins for Deciphering Functions. *Chemistry & Biology.* 2014;21(1):51-66.

126. Grogan MJ, Pratt MR, Marcaurelle LA, Bertozzi CR. Homogeneous glycopeptides and glycoproteins for biological investigation. *Annu Rev Biochem.* 2002;71:593-634.

127. Guo Z, Shao N. Glycopeptide and glycoprotein synthesis involving unprotected carbohydrate building blocks. *Med Res Rev.* 2005;25(6):655-78.

128. Hojo H, Nakahara Y. Recent progress in the field of glycopeptide synthesis. *Biopolymers.* 2007;88(2):308-24.

129. Calin O, Eller S, Seeberger PH. Automated polysaccharide synthesis: assembly of a 30mer mannoside. *Angew Chem Int Ed Engl.* 2013;52(22):5862-5.

130. Weishaupt MW, Matthies S, Seeberger PH. Automated solid-phase synthesis of a beta-(1,3)-glucan dodecasaccharide. *Chemistry.* 2013;19(37):12497-503.

131. Esposito D, Hurevich M, Castagner B, Wang CC, Seeberger PH. Automated synthesis of sialylated oligosaccharides. *Beilstein J Org Chem*. 2012;8:1601-9.
132. Wu B, Hua Z, Warren JD, Ranganathan K, Wan Q, Chen G, et al. Synthesis of the fucosylated biantennary N-glycan of erythropoietin. *Tetrahedron Lett*. 2006;47(31):5577-9.
133. Kajihara Y, Suzuki Y, Yamamoto N, Sasaki K, Sakakibara T, Juneja LR. Prompt Chemoenzymatic Synthesis of Diverse Complex-Type Oligosaccharides and Its Application to the Solid-Phase Synthesis of a Glycopeptide with Asn-Linked Sialyl-undeca- and Asialo-nonasaccharides. *Chemistry – A European Journal*. 2004;10(4):971-85.
134. Kajihara Y, Yamamoto N, Miyazaki T, Sato H. Synthesis of diverse asparagine linked oligosaccharides and synthesis of sialylglycopeptide on solid phase. *Curr Med Chem*. 2005;12(5):527-50.
135. Yamamoto N, Ohmori Y, Sakakibara T, Sasaki K, Juneja LR, Kajihara Y. Solid-Phase Synthesis of Sialylglycopeptides through Selective Esterification of the Sialic Acid Residues of an Asn-Linked Complex-Type Sialyloligosaccharide. *Angew Chem Int Ed Engl*. 2003;42(22):2537-40.
136. Piontek C, Ring P, Harjes O, Heinlein C, Mezzato S, Lombana N, et al. Semisynthesis of a homogeneous glycoprotein enzyme: ribonuclease C: part 1. *Angew Chem Int Ed Engl*. 2009;48(11):1936-40.
137. Piontek C, Varon Silva D, Heinlein C, Pohner C, Mezzato S, Ring P, et al. Semisynthesis of a homogeneous glycoprotein enzyme: ribonuclease C: part 2. *Angew Chem Int Ed Engl*. 2009;48(11):1941-5.

2. IN-DEPTH STRUCTURAL CHARACTERISATION OF N- GLYCANS USING PGC-NANO-LC-ESI-MS/MS

Synopsis

The primary in-depth glycomics tool presented in this work is based on a Porous Graphitized Carbon Liquid Chromatography (PGC-LC) ElectroSpray Ionisation (ESI) tandem Mass Spectrometry (MS/MS) glycomics approach.

This chapter has been modified in part from the following articles:

Alagesan K, Möglinger U, Altmann F & Kolarich D. (2016). 'Better together – relative retention time plus spectral matching improves automated glycan characterisation using PGC-nLC-IT-ESI-MSMS' – *Manuscript in preparation*

Campbell MP, Nguyen-Khuong T, Hayes CA, Flowers SA, **Alagesan K**, Kolarich D, Packer NH, Karlsson NG. (2014). 'UniCarb-DB: A curation pipeline to build a global glycan reference MS/MS repository' – *Biochim. Biophys. Acta.*c 1844(1 Pt A):108-16.

Kolarich D, Windwarder M, **Alagesan K** & Altmann F. (2015). 'Isomer specific analysis of released N-glycans by LC-ESI MS/MS with porous graphitized carbon' – *Methods in Molecular Biology*; 1321: 427-435.

2.1 BACKGROUND

The intrinsic complexity of glycan sequences (non-linearity, various types of linkages possible within two building blocks, non-template based biosynthesis) requires multi-dimensional techniques for glycome sequencing to accommodate both, in-depth structural data acquisition and high throughput (1). Mass spectrometry (in particular in combination with orthogonal techniques such as LC) provides the potential capability to fully define the composition, sequence and topology of complex glycans.

Structure resolved glycan mass profiling by MALDI-TOF MS necessitates the derivatisation of released glycans. In most cases, permethylation of released glycan is preferred as it improves the ionisation efficiency and also stabilises labile substituents such as sialic acids. Nevertheless, glycan permethylation might also introduce unintended artefacts if not performed properly (2). Robinson *et al.* observed a series of artefact ions that were all +30 Da larger than the completely permethylated glycans due to the incorporation of a methoxymethyl instead of methyl group. This phenomenon can in particular result in misinterpretations since the mass difference between permethylated Fucose and Hexose residues as well as between permethylated NeuAc and NeuGc also corresponds to 30 Da (3). On contrast, ESI-MS/MS in combination with suitable online liquid chromatography separation provides the additional advantage that isobaric glycan structures, which are not distinguishable by mass alone, can be separated in the chromatographic dimension.

In recent years, PGC has emerged as a robust and versatile chromatographic medium for LC-separation of non-derivatised oligosaccharides due to its remarkable ability to separate isobaric glycans. PGC-LC-ESI-MS/MS in negative ion mode allows for an unambiguous and detailed glycan structure characterisation of isobaric *N*- and *O*-glycans within a single LC-MS/MS experiment, drastically reducing the sample amount (Figure 2.1). The combination of PGC-LC with ESI ion trap MS/MS provides a powerful platform to acquire distinct and specific tandem mass spectra for isobaric glycans which otherwise cannot be separated and individually identified by conventional mass spectrometric techniques. This capacity, however, is crucial for the in depth glycan sequencing of individual proteins and entire glycomes (4).

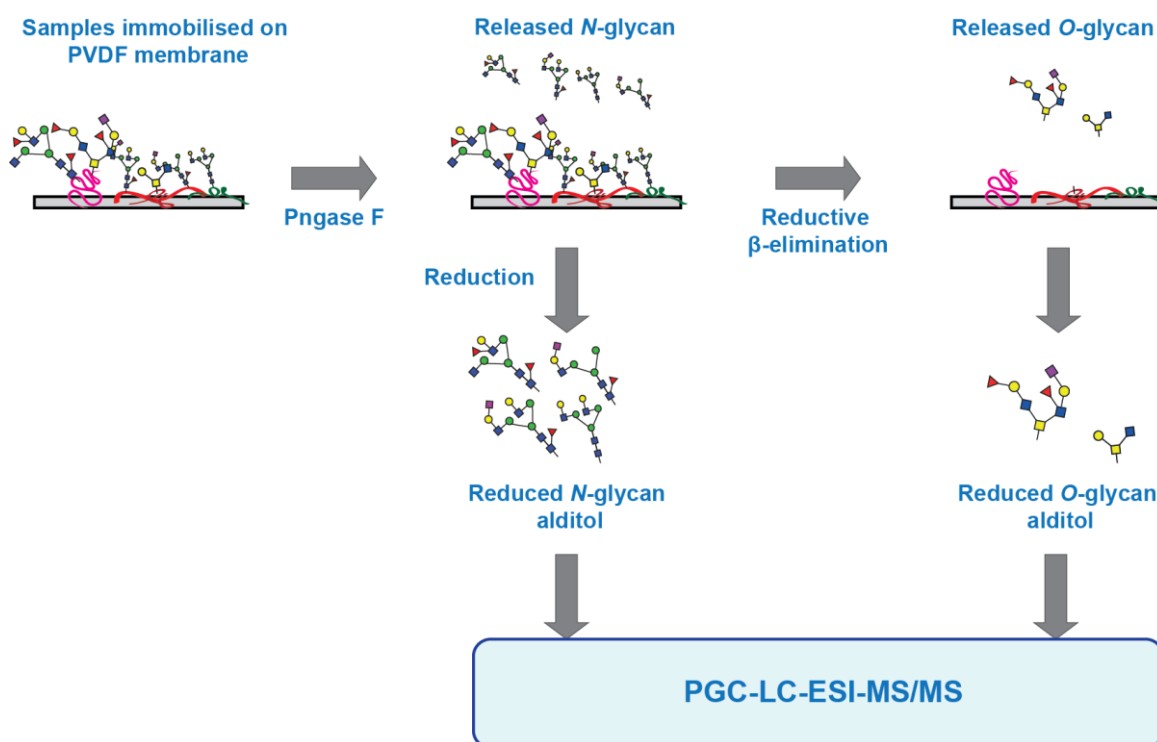


Figure 2.1: Generalised workflow for PGC-LC ESI MS/MS based N- and O- glycomics using the same sample.









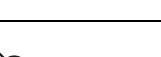
Interpretation of glycan fragmentation data is often time consuming and represents a major limiting factor for high-throughput MS/MS based glycomics. Automated spectral matching and scoring provides a better alternative to the laborious and repetitive manual annotation of the tandem MS spectra. Establishment and collection of well annotated glycan spectral libraries thus presents one opportunity with a huge potential to facilitate reliable structure assignment supporting data analysis automation. One part of the puzzle towards automated PGC-LC-ESI MS/MS based glycomics is a reliable and robust workflow that allows the creation of a well curated spectral library starting with commercially available N-glycan alditol standards with known composition, linkage and branching information. These standard glycan spectral libraries have been developed in partnership with UniCarb-DB (5). The pilot study focused on a comparative analysis of MS/MS fragmentation patterns acquired by different ion trap instruments using similar protocols and chromatographic conditions. This comprehensive analysis of a wide range of oligosaccharide standards provided valuable information about glycan fragmentation and was subsequently used to develop a fully customisable spectral matching and scoring tool incorporating relative retention time (RRT) for automated glycan identification.

2.2. MATERIALS AND METHODS

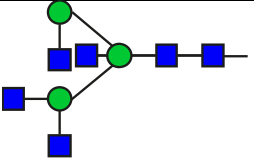
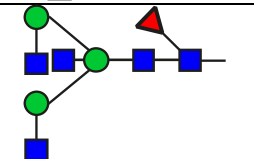
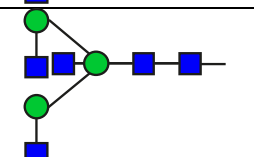
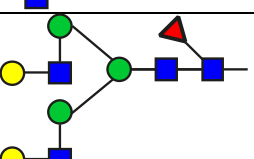
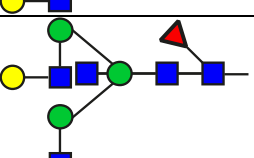
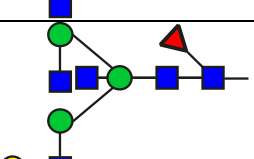
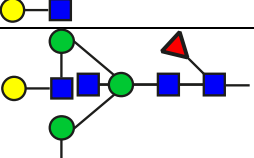
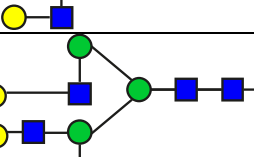
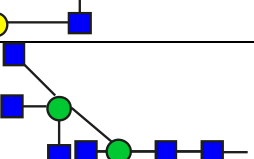
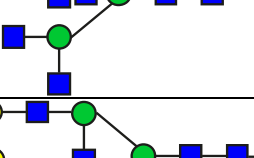
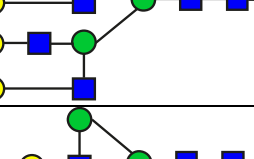
2.2.1. GLYCAN STANDARDS

Oligosaccharide standards were supplied in kind by Professor Jeremy Turnbull, University of Liverpool and Professor Friedrich Altmann, University of Natural Resources and Applied Life Sciences, Vienna (Table 2.1).

Table 2.1: Glycan standards used to compare negative ion mode fragmentation of glycan and to develop Glyco-RRT-MS/MS library.

SNo	Oligosaccharide standard	Proglycan Name	Precursor mass	
			$[M-2H]^{2-}$	$[M-H]^{-}$
1		M1	293.11	587.23
2		M1F	366.14	733.30
3		M2	374.13	749.31
4		M3	455.16	911.37
5		M4	536.19	1073.38
6		M4	536.19	1073.38
7		M5	617.21	1235.57
8		M6	698.24	1397.49
9		M9	941.32	1883.65

10		M9 Glu1	1022.34	2045.69
11		GnGnF	731.27	1463.55
12		GnMan1F	556.74	1114.46
13		Man1GnF	556.74	1114.46
14		Man1Man1F	528.19	1057.39
15		GnGn	658.24	1317.49
16		Gn(GnGn)	759.78	1520.57
17		AGn	739.27	1479.54
18		GnA	739.27	1479.54
19		AA	820.29	1641.60
20		(GnGn)(GnGn)	861.32	1723.63

21		Gn(GnGn)bi	861.32	1723.63
22		GnGnFbi	832.81	1666.63
23		GnGnbi	759.78	1520.57
24		AAF	893.30	1787.65
25		AGnFbi	913.83	1828.68
26		GnAFbi	913.83	1828.68
27		AAFbi	994.86	1990.73
28		A(AA)	1002.86	2006.73
29		(GnGnGn) (GnGn)bi	1064.40	2129.81
30		(AA)(AA)	1185.42	2371.86
31		Na(2-3)A	965.84	1932.69

32		ANa(2-3)	965.84	1932.69
33		Na(2-3) Na(2-3)	1114.00	2223.79
34		A(AF)	893.37	1787.65
35		A(AF)	893.37	1787.65
36		(AF)(AF)	966.35	1933.71
37		Na(2-6) Na(2-6)	1111.40	2223.79
38		ANa(2-6)	965.84	1932.69
39		Na(2-6) A	965.84	1932.69
40		Man3Gnbi	820.34	1438.52

2.2.2. SDS-PAGE AND ELECTRO-BLOTTING

Five μg of protein (2.5 μL in PBS) were incubated with 4 μL 4x SDS-PAGE sample buffer (0.25 M Tris/HCl pH 6.8, 40% [v/v] Glycerin, 4% [w/v] SDS, 0.015% (w/v) Bromophenol blue) containing 50 mM dithiothreitol (DTT) at 96°C for 5 minutes. After cooling down to room temperature, an aliquot of iodoacetamide solution (500 mM) was added to a final concentration of 50 mM iodoacetamide and incubated for 30 minutes in the dark before the proteins were subjected to SDS-PAGE separation on a precast 10% Mini-PROTEAN® TGX™ Gel (Biorad). Electrophoresis was performed at 200 V until the indicator band reached to bottom end of the gel. After SDS-PAGE, proteins were subsequently semidry electroblotted onto PVDF membranes (0.2 μm pore size, Biorad, Munich, Germany) using the Trans-Blot® Turbo™ Transfer System (Biorad). Blotting was performed according to manufacturer's recommendations (7 min at 2.5 A and 25 V). Transferred proteins were visualised after staining with direct blue 71 as described previously (6).

2.2.3. GLYCAN RELEASE

Direct blue 71 stained bands were cut from the PVDF membrane and transferred into 96 well plates for sequential *N*- and *O*-glycan release from the PVDF membrane as described previously (6). Any free PVDF surface was blocked with polyvinylpyrrolidone solution (1% PVP 40 in 50% methanol) and incubated over night at 37°C using 15 μL *N*-Glycosidase F (PNGase F, Roche Diagnostics, Mannheim, Germany) solution (0.17 U/ μL in 10 mM NH_4HCO_3). *N*-glycans were collected and reduced for 3 hours in 1 M NaBH_4 in 50 mM KOH. The *N*-glycan free sample on the PVDF membrane was subjected to a chemical *O*-glycan release via reductive beta-elimination (0.5 M NaBH_4 in 50 mM KOH, 50°C for 16 hours) and collected from the 96-well plates as described previously (6, 7). All samples were subsequently desalted by cation exchange chromatography (Dowex 50wX2, Biorad) using self-made micro spin columns as described in detail previously. After desalting, samples were further subjected to a carbon clean up via PGC micro-spin columns.

2.2.3. PGC CLEAN UP

A filter TopTip (Glygen, Columbia, MD) was filled with PGC material (resin was obtained from an Alltech Extract-Clean™ Carbograph SPE Column, Deerfield, IL).

The micro spin column was washed three times with 50 μ L 80% acetonitrile containing 0.1% TFA and subsequently equilibrated by washing three times with 50 μ L 0.1% TFA. The samples were loaded and the columns washed two times with 50 μ L 0.1% TFA before glycans were eluted using 2x50 μ L 80% acetonitrile containing 0.1% TFA. Eluted glycan samples were dried in the Speed Vac concentrator without any additional heat. Glycan samples were then reconstituted in 20 μ L water for PGC nano liquid chromatography - ESI MS/MS analysis, a 3 μ L aliquot was injected.

2.2.4. PGC NANO LC-ESI IT-MS/MS ANALYSIS

PGC-LC was performed on an Ultimate 3000 UHPLC system (Dionex, Part of Thermo Fisher, Germany) online coupled to an amaZon speed ETD ion trap mass spectrometer (Bruker Daltonics, Bremen, Germany). The LC system was connected to the mass spectrometer using the nano-flow ESI sprayer (Bruker Daltonics, Bremen, Germany). The instrument was controlled using Hystar 3.2 software. The spectra were analysed using Compass Data Analysis v4.2 (Bruker Daltonics, Bremen). The instrument was set up to perform CID fragmentation on the selected precursors. An m/z range from 350-1600 Da was used for data dependant precursor scanning. The three most intense signals in every MS scan were selected for MS/MS experiments. MS as well as MS/MS data were recorded in the instrument's "Ultra scan mode" and "Enhanced resolution mode" respectively. Specific instrumental operational parameters used in the present investigation are listed in Table 2.2. Glycans were loaded onto a PGC (porous graphitized carbon) precolumn (Hypercarb KAPPA 30 x 0.32 mm, 5 μ m particle size) and separated on an analytical PGC column (Hypercarb™ PGC Column, 100 mm x 75 μ m particle size 3 μ m, both ThermoFisher Scientific, Waltham, MA). The samples were loaded onto the precolumn at a flow rate of 6 μ L/min in 98% buffer A (10 mM ammonium bicarbonate). The starting conditions for the analytical column at a flow rate of 1 μ L/min were 3% buffer B (10 mM ammonium bicarbonate in 60% acetonitrile). The gradient conditions were as follows: increase of buffer B from 3 to 16% (5.5-7.5 min), further increase to 40.0% B (7.5-55.5 min), followed by a steeper increase to 90% B (55.5-62.0 min). The column was held at 95% B for 8 min (62-70 min). At the same time the precolumn was flushed with 90% Buffer C (10 mM ammonium bicarbonate in 90% acetonitrile) at a flow rate of

6 $\mu\text{L}/\text{min}$ before reequilibrating the precolumn as well as the analytical column in 98% buffer A for 7 minutes. All analyses were performed in triplicate.

Table 2.2: Ion-trap MS settings for glycan analysis by PGC-nano-LC-ESI-MS/MS

Parameter	negative mode / nano
MS Settings - General	
Capillary voltage	1-1.3 kV
SPS	m/z 900
Compound stability	100%
Trap Drive Level	100%
Spectra averaging	5
Dry gas temperature	150°C
Dry Gas flow	3 L/min
Maximum accumulation time	200 ms
Ion mode	negative
MS-Scan	
MS Scan mode	ultrascan
ICC target	40000
Mass detection range	m/z 350-1600
MS2	
MS scan mode	Enhanced
SPS MS(n)	automatic
MS(n) spectra averages	5
MS(n) ICC target	150000
Preferred charge state	None
Active exclusion	Off
Mass detection range	m/z 100-2500
Isolation width	3 Da
Exclude singly charged ions	off
No. of precursor ions	3
SmartFrag	Enhanced
SmartFrag Start amplitude	30%
SmartFrag End amplitude	120%
Fragmentation width	5 m/z

2.2.5. SPECTRAL LIBRARY DESIGN AND IMPLEMENTATION

The library was created using LibraryEditor V4.2, which is a part of Compass Data Analysis v4.2 (Bruker Daltonics, Bremen) and Excel (Microsoft). The library features meta data features such as absolute chromatographic retention time of the reduced *N*-glycan alditols, mass (*m/z*) and tandem MS/MS spectra.

2.2.5.1. LC-MS and MS/MS Data

The extracted ion chromatogram (EIC) for each glycan was derived based on the glycan $[M-H]^-/[M-2H]^{2-}$ masses. Absolute retention times of the EIC of the given *m/z* values under the experimental conditions were recorded and exported to the library.

Relative Retention time (RRT) was calculated as the ratio of absolute retention time of the glycan to the absolute retention time of the query.

2.2.5.2. MS/MS Spectral Library

To simplify the product ion spectra obtained for each standard *N*-glycan, peaks in each MS/MS spectra were merged with a mass tolerance for the precursor (0.5 Da). The diagnostic ions observed in the negative mode fragmentation were also incorporated into the library.

2.2.5.3. Spectral matching and scoring

The dot-product function available in the R package OrgMassSpecR (<http://orgmassspecr.r-forge.r-project.org>), a general package for mass spectrometry analysis, was used to calculate the similarity scores of the isobaric *N*-glycan structures.

2.2.5.4. Library Evaluation by Spectral Searching

Before implementing the GlycanRRT library in the glycan analysis workflow, the datasets used to create the library was re-searched the GlycanRRT library.

2.2.5.5. LC-MS/MS blind trial

Three different LC-MS/MS glycan datasets from different participants were selected for the blind trial. The blind trial spectra were processed and then searched against all the spectra present the GlycanRRT library. The results were exported into excel containing all the meta data including % fit score.

2.2.6. GLYCAN IDENTIFICATION AND RELATIVE QUANTITATION

Glycans were automatically identified using an in house established GlycoRRT spectral database and the spectral library tool integrated in Compass Data Analysis 4.2 (Bruker). *N*-glycan structures not present in the database were manually annotated using Data Analysis 4.2. After identification of the *N*-glycans, relative quantitation was performed using the QuantAnalysis tool (Bruker), which determines the area under the curve obtained from the individual extracted ion chromatograms (EIC) from multiple analyses. The integration of every extracted ion chromatogram was validated manually, in particular for peaks with very low abundance. The values out of three technical replicates were averaged for this study and their standard deviations are represented in the error bars. In each sample the total amount of the identified glycan structures were taken as 100% and their relative distribution determined.

2.3. RESULTS AND DISCUSSION

2.3.1. RATIONALE AND LIBRARY DESIGN

PGC is gradually recognised as a highly useful stationary phase for the online separation of reduced *N*- and *O*-glycan alditol glycans as part of a negative ion mode LC-ESI-MS/MS glycomics workflow due to its high separation capacity allowing the separation of various glycan isomers. However, the major bottle-neck for such a high throughput glycomics technique remains the automated data analysis due to lack of reliable software tools. A MS/MS spectra matching approach querying acquired spectra against well defined, library ones will facilitate reliable such a structural assignment, subject to a full understanding of all analytical parameters that influence retention time, signal intensity and resulting MS/MS spectra data. Thus as part of this thesis a fully customisable and easily adaptable tool for automated glycan identification combining spectral matching with relative retention time has been developed. Various parameters influencing glycan spectral matching and PGC retention were systematically evaluated including the influence of instrumental parameters on glycan quantification. The developed Glycan RRT LC–MS/MS Library contains critical information pertaining to MS instrumentation, experimental conditions, retention times, fragmentation profiles and/or structural diagnostic ion features (Figure 2.2).

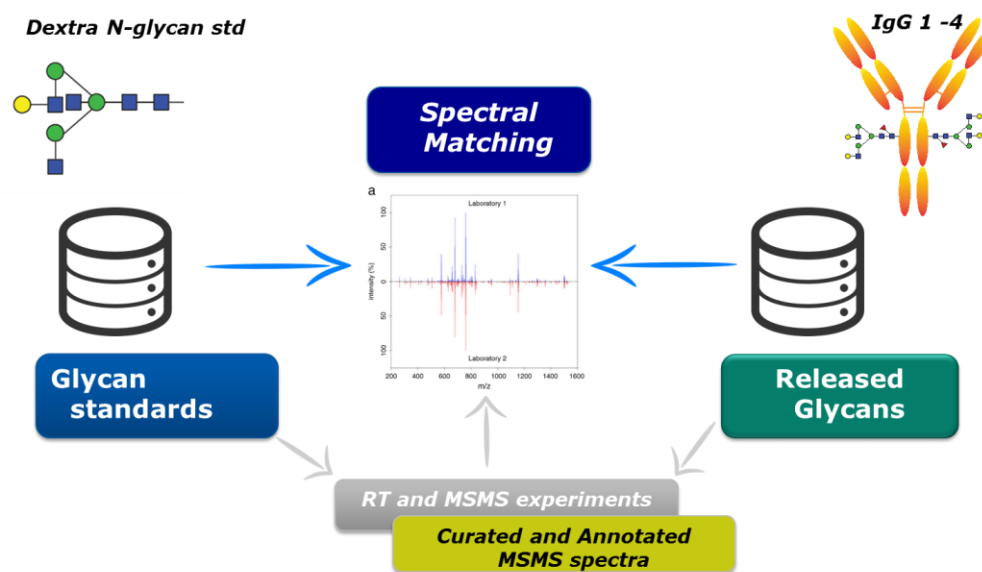


Figure 2.2: Glycan Spectral library design. The Glycan RRT LC–MS/MS Library contains carefully annotated negative ion ESI-MS fragmentation spectra of over 200 N-glycan structures together with MS and experimental meta data. Automated glycan identification is facilitated by spectral matching of query spectra against well annotated reference spectra and the spectral scoring is provided as %fit score.

2.3.2. NEGATIVE ION MODE PGC-LC-ESI-MS/MS GLYCOMICS

Fragment ions generated in the negative-ion MS/MS spectra are rich in information that assists meticulous glycan structural assignment. In contrast to positive ion fragmentation, in negative ion fragmentation diagnostic ions are produced by a single pathway following proton abstraction from the specific hydroxyl groups (8). Thus the observed diagnostic fragment ions derived from negative ion MS/MS spectra can be used to elucidate structural features such as the position of fucose, the core type of O-glycans and the branching of N-glycan structures (9-11). The presence of an abundant [D-221] ion (e.g., $m/z = 508$ or 670 or 961), for example, indicates the presence of a bisecting GlcNAc residue whereas as the Z1 ($m/z = 350$) and Z2 ($m/z = 553$) diagnostic ions are indicative for the presence of core fucosylation (Figure 2.3)

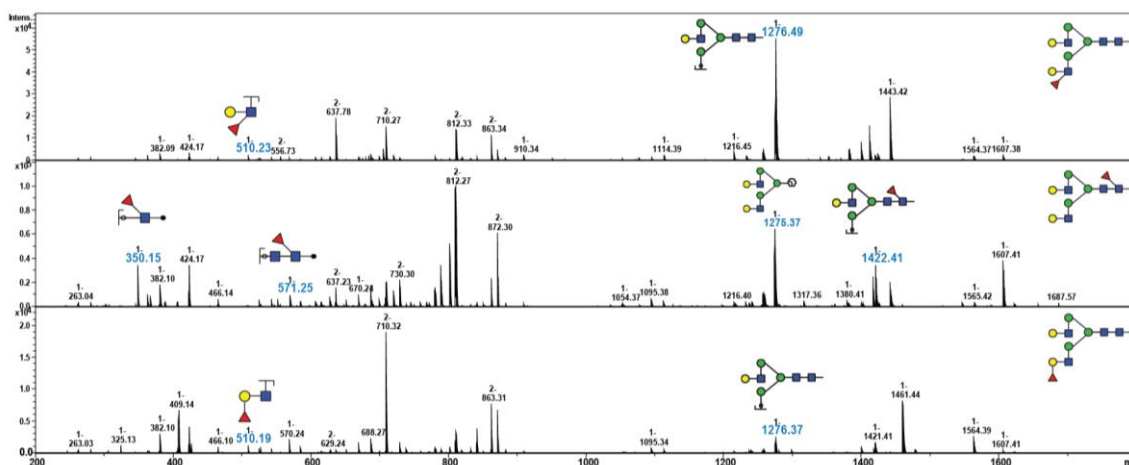


Figure 2.3: Characterisation of neutral fucosylated *N*-glycan structural features by negative ion PGC-nLC-ESI-IT-MS/MS. The presence of core fucose is indicated by Z1 ion at *m/z* 350 and Z2 ion at 571, whereas the presence of arm fucosylation can be differentiated by B and C ions at *m/z* 510 and 528.

Specific glycan structure features are known to influence the PGC elution behaviour. The order of elution of individual structural isomers is also highly conserved (12). *N*-glycans carrying a bisecting *N*-acetylglucosamine (GlcNAc) elute earlier than their non-bisected structural isomers. Similarly, presence of core fucosylation increases retention whereas Lewis type antenna-fucosylation results in shorter retention times. The polar retention effect accounts for increased selectivity, in particularly for sialylated glycans. Very subtle changes such as the linkage position of *N*-acetylneuraminic acid residues significantly alter the retention time of isobaric glycan structures. The glycan structures carrying α 2,3-linked NeuAc residues are retained much stronger compared to their α 2,6-linked counterparts. In addition to their different elution time, the in negative mode MS/MS spectra of the PGC separated compounds also provide additional confirmation on the NeuAc linkage positions (Figure 2.4). The high separation power of PGC-LC is just not limited to sialylated residues but also equally applies to other *N*-glycan classes such as high mannose isomers and complex *N*-glycans structures carrying Lewis x/a/y/b structures. Thus combining the high separation capacity offered by PGC with information rich negative mode fragmentation represents a unique technology that acquires multidimensional data that all together provides profound structural and quantitative information also from very complex sample sources while using a minimum of initial sample – a crucial prerequisite for any glycan sequencing technology that is to be applicable for clinical research.

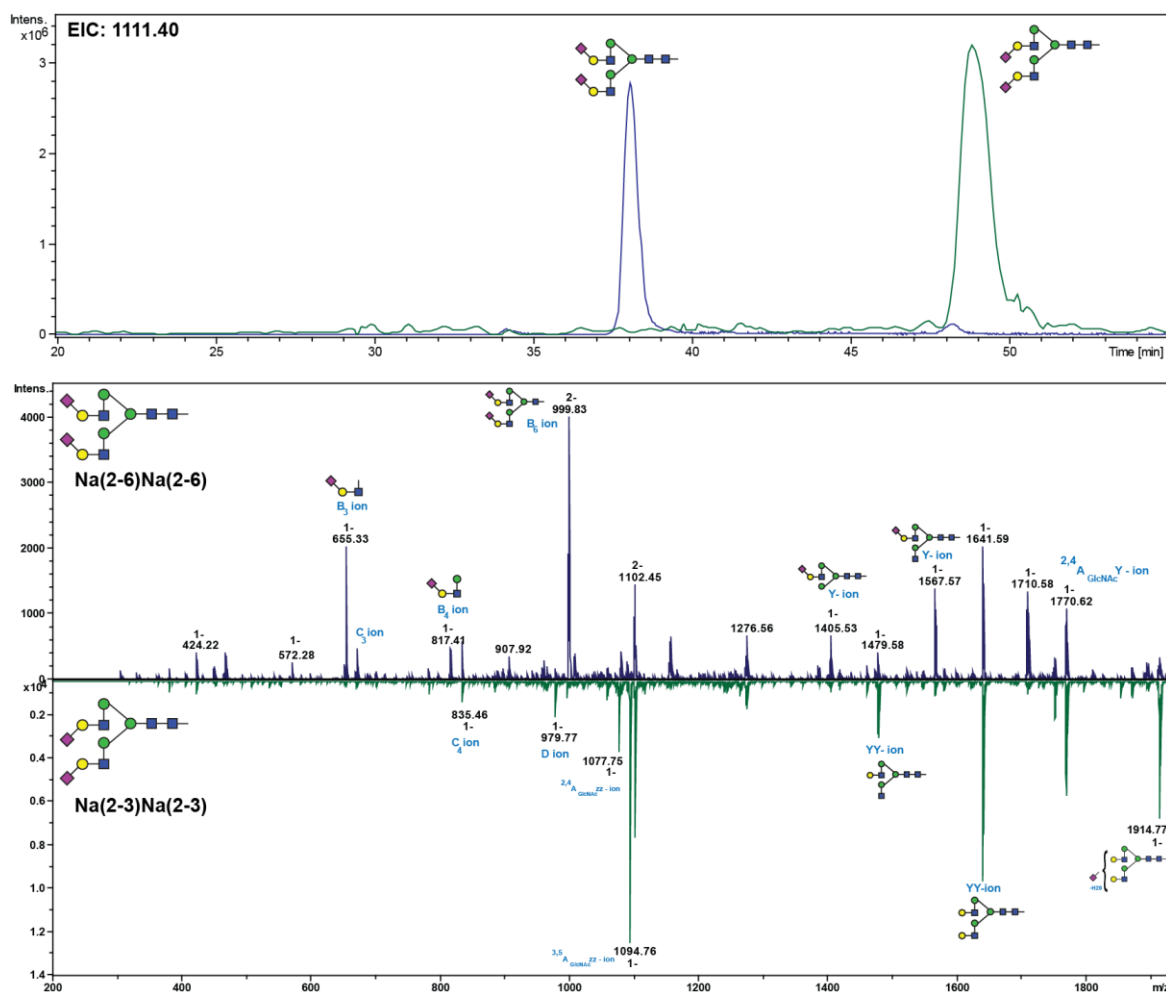


Figure 2.4: PGC-LC separation of α 2,6 and α 2,3 linked Neu5Ac isomers. Extracted ion chromatogram (EIC) of $[M-2H]^{2-} = 1111.40$ corresponding to the composition $\text{Hex}_2\text{HexNAc}_2\text{NeuAc}_2 + \text{Man}_3\text{GlcNAc}_2$ eluting at two different time points (Top panel). In addition to different elution profiles, the respective negative mode MS/MS spectra showed structure specific MS/MS signatures that allowed isomer differentiation. The prominent water loss (-18 Da) observed for the α 2,3 linked Neu5Ac (m/z 1914.77) indicated the presence of a α 2,3 linked Neu5Ac (bottom panel).

2.3.3. AN ASSESSMENT OF SPECTRAL QUALITY AND SIMILARITY SCORING

The spectral similarity between the reference and the query spectra were expressed in a similarity score calculated using the dot-product function available in the R package OrgMassSpecR. Previous work performed on the MS analysis of small molecules has demonstrated that the spectral dot product is the most effective scoring function (13). In spectral library matching the ideal situation is to obtain a perfect match of the unknown experimentally derived spectrum with a single, well annotated library spectrum with a one-to-one correspondence between each peak intensity and m/z value. A value of 1 indicates a perfect overlay of the

two MS/MS spectra. In reality however, because of unavoidable instrument variability this is theoretical value of 1 is unlikely to be achieved.

The principle objective of the initial multi-institutional study was to compare the acquired negative mode MS/MS fragmentation data from different instruments and laboratories and to identify data parameters that influence the similarity scores of the compared spectra. To this end, the reproducibility of product ion spectra obtained from three separate laboratories on the same set of 23 standard oligosaccharides was assessed. For each set of spectra acquired from the three ion trap instruments, a similarity score was calculated to determine spectral similarity, using a head-to-tail plot of two mass spectra with the query spectrum (bottom) and a reference spectra (top) (Table 2.3).

An example alignment for the asialo, agalacto and tetra-antennary glycans is shown in Figure 2.5. The laboratory 1 spectra were used as the reference spectra. Here, we observed that tandem spectral comparisons between 0.92 and 0.96 have a degree of similarity between the most intense ions in relation to intensity and m/z . The scores in Table 2.3 ranged between 0.5 [(AA)(AA)] to 0.99 [Na(2-6) Na(2-6)] for laboratory 2 and 0.01 [AA] to 0.99 [Na(2-6) Na(2-6)] for laboratory 3 with the performance for the optimized dot-product approach on average 0.82 and 0.77 for glycans analysed by laboratory 2 and laboratory 3 relative to the reference laboratory (laboratory 1). From these analyses it becomes evident that the fragmentation similarity is high in most cases and the assigned statistical confidence consolidates the reproducibility between the institutes, despite minor differences in sample preparation and data acquisition methods.

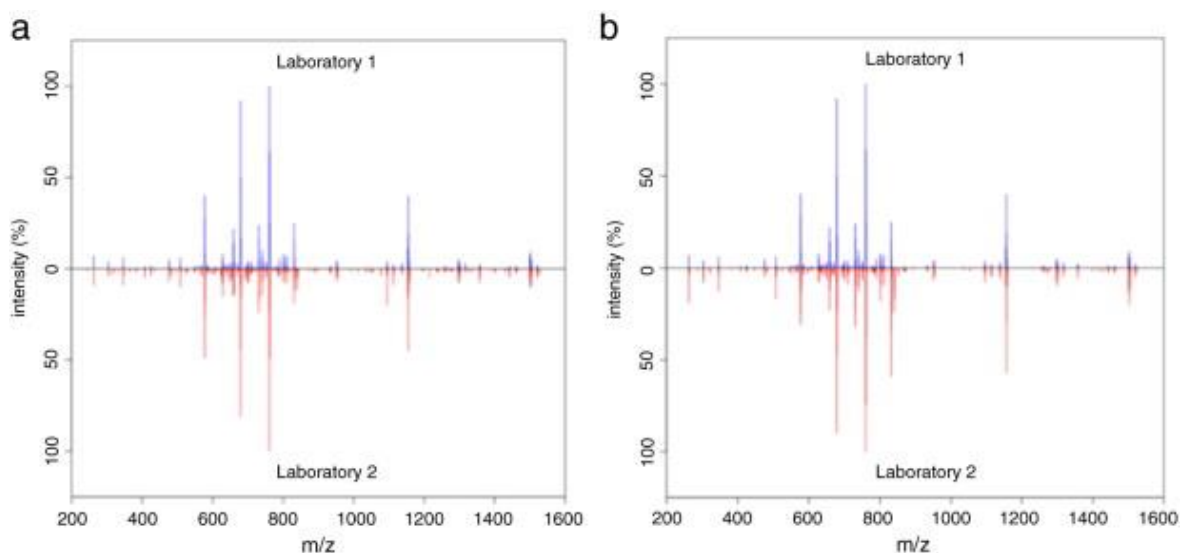


Figure 2.5: Representative MS/MS spectra acquired for the asialo, agalacto and tetraantennary [(GnGn)(GnGn)] glycans standard analysed at laboratory 1 (top spectra in blue) compared to (a) laboratory 2 with a 0.96 similarity score, and (b) laboratory 3 with a 0.92 similarity score. (Figure taken from (14))

Table 2.3: Similarity scores for each glycan standard. The ion trap mass spectrometers used in this study were Thermo Finnigan LTQ linear ion trap (laboratory 1), Agilent XCT-ultra ion trap (laboratory 2) and Bruker amaZon ETD speed ion trap (laboratory 3).

Glycan	Laboratory 2	Laboratory 3
Gn(GnGn)	0.56	0.97
GnGnF	0.84	0.53
GnA	0.69	0.18
AA	0.91	0.01
(GnGn)(GnGn)	0.96	0.92
Gn(GnGn)bi	0.92	0.95
GnGnFbi	0.68	0.96
AAF	0.67	0.68
AGnFbi	0.88	0.5
A(AA)	0.81	0.87
(GnGnGn) (GnGn)bi	0.68	0.8
AAFbi	0.87	0.52
(AA)(AA)	0.5	0.76
M1	0.67	0.98
M1F	0.96	0.98
M2	0.92	0.98
M3	0.94	0.93
Man1Man1F	0.94	0.97
M5	0.77	0.94

M6	0.97	0.75
M9Glc1	0.85	0.91
Na(2-6) A	0.85	0.59
Na(2-6) Na(2-6)	0.99	0.99

2.3.4. INFLUENCE OF THE CHARGE STATE ON SPECTRAL ALIGNMENT

Two significant outliers were reported for the sets AA and GnA from laboratory 3. Such low scores are attributed to the comparison of MS/MS spectra with different precursor m/z ratios *i.e* comparing fragmentation of $[M-H]^-$ with $[M-2H]^{2-}$. **Figure 2.6-A** shows the MS/MS spectra of standard GnA from laboratories 1 and 3. These data showed that the doubly charged species were more prone to glycosidic cleavages, in particular Y fragments, whereas in the singly charged species X-type cross-ring cleavage (loss of $C_2H_4O_2$) products were among the most intense peaks. The X-type cross-ring fragmentation is responsible for the prevalence of satellite peaks below the Y and Z fragments of the singly charged species, especially in the m/z 1300 to 1500 region. This was in contrast to the low m/z region of the doubly charged species where more A-type cross-ring fragments were detected, providing quality sequence information as opposed to the inconclusive X-type fragmentation (**Figure 2.6-B**).

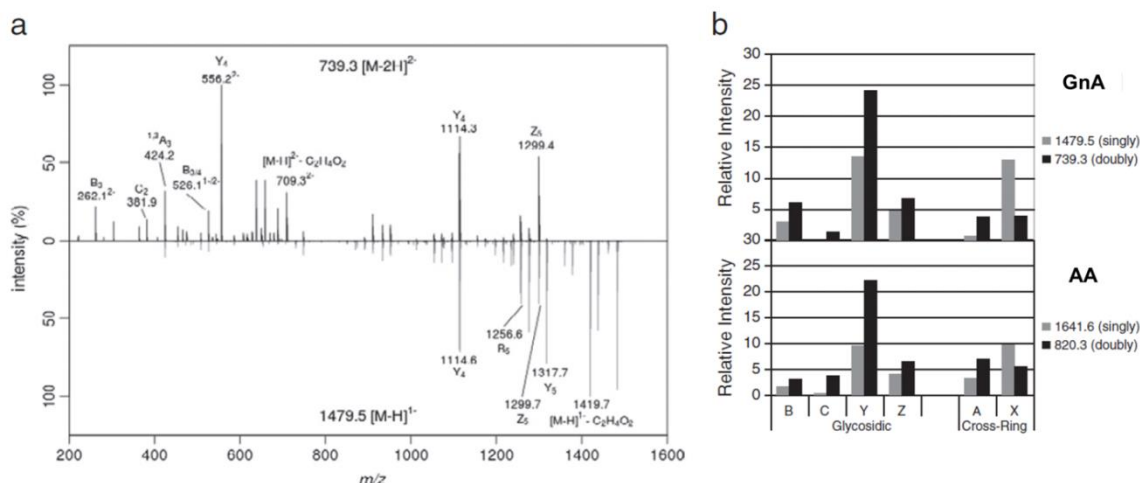


Figure 2.6: (a) A comparison of the MS/MS fragmentation of the GnA glycan standard acquired from the singly and doubly charged precursors by laboratories 1 and 3, respectively. Bar charts (b) show the relative intensities of selected fragment ions observed for GnA (upper) and AA (lower) [figure taken from (14)].

Next, the reproducibility of the MS/MS spectra within the same instrument over a five-year period (2012 – 2016) was investigated. The MS/MS spectra recorded in 2016 served as the reference spectra for similarity score calculation and the various fragment spectra recorded for the very same glycan structure have been compared (Figure 2.7: A-D) The similarity score for various fucosylated *N*-glycan is shown in the Table 2

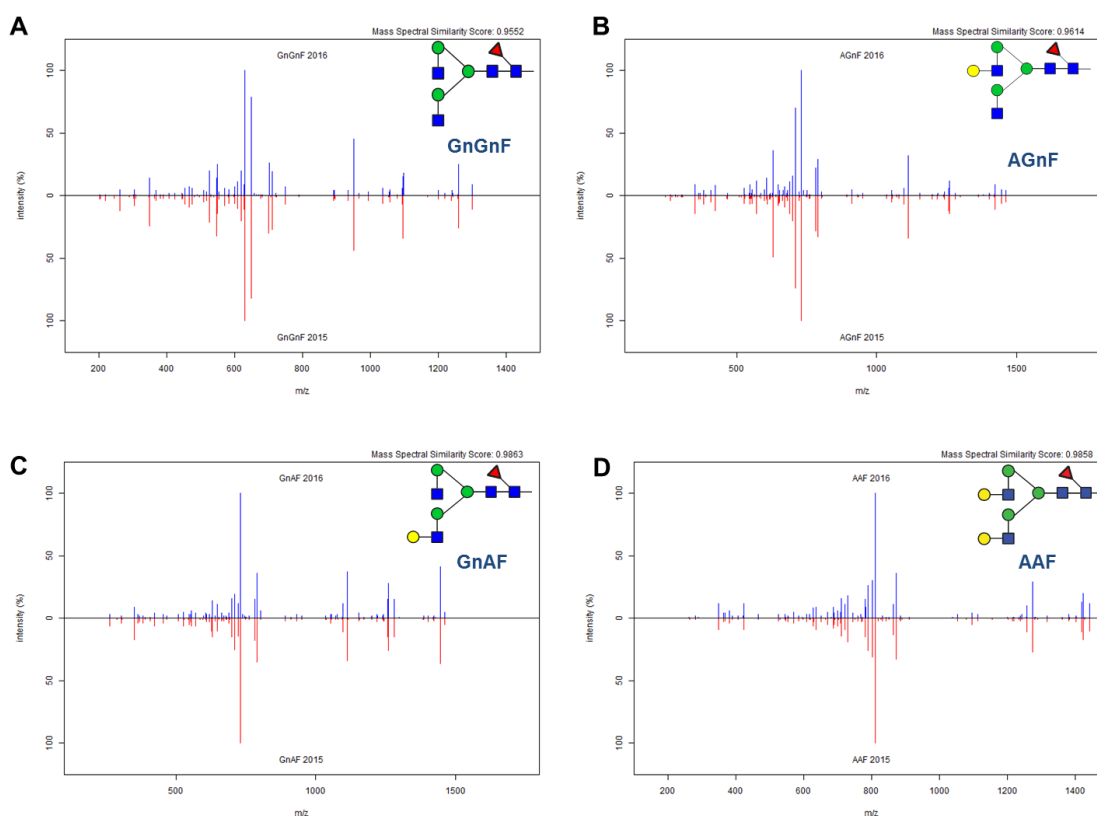


Figure 2.7: Spectra alignment of tandem MS spectra from four exemplary neutral fucosylated *N*-glycans (A) GnGnF (B) AGnF (C) GnAF (D) AAF. Bottom spectra in red were acquired in 2015 and have been compared to the reference spectra of the same *N*-glycans recorded in 2016 (top blue spectra).

Table 2.4: Similarity scores for various neutral fucosylated N-glycans obtained over the period of 5 years.

Similarity Score				
Year	GnGnF	AGnF	GnAF	AAF
2012	0.89	0.91	0.91	0.93
2013	0.94	0.95	0.97	0.98
2014	0.98	0.98	0.98	0.98
2015	0.96	0.96	0.99	0.99
2016	1.0	1.0	1.0	1.0
SD	0.04	0.03	0.03	0.03

2.3.5. FACTORS INFLUENCING GLYCAN SEPARATION BY PGC-LC

Although the exact molecular nature of the retention mechanism leading to isobaric glycan separation by PGC-LC is yet still not fully understood, hydrophobic, polar and ionic interactions have been identified to contribute to analyte retention (15, 16). Unlike hydrophilic-interaction liquid chromatography (HILIC) and ion pairing reversed-phase chromatography (IP-RPC), glycan retention in PGC cannot be correlated to the monosaccharide compositions (hydrophilicity of the glycan) by multiple linear regression analysis (17). On the other hand, factors such as electrosorption, solvent, temperature and ion polarity are also known to influence the separation behaviour of sialylated glycans (18). Yet, little is known about the influence these parameters have on the isomer separation capacity offered by PGC. Thus the influence the column temperature has on the isomer separation capacity was evaluated using AGnF and GnAF N-glycans. In agreement with previous studies(18), higher column temperatures did result in an increased glycan retention capacity (Figure 2.8). Besides, the here presented data also indicates that an increase in column temperature also resulted in better baseline separation of the AGnF and GnAF isomers as evident from the m/z 812.33 extracted ion chromatogram (EIC) corresponding to doubly charged signal of these two standard glycans (Figure 2.8).

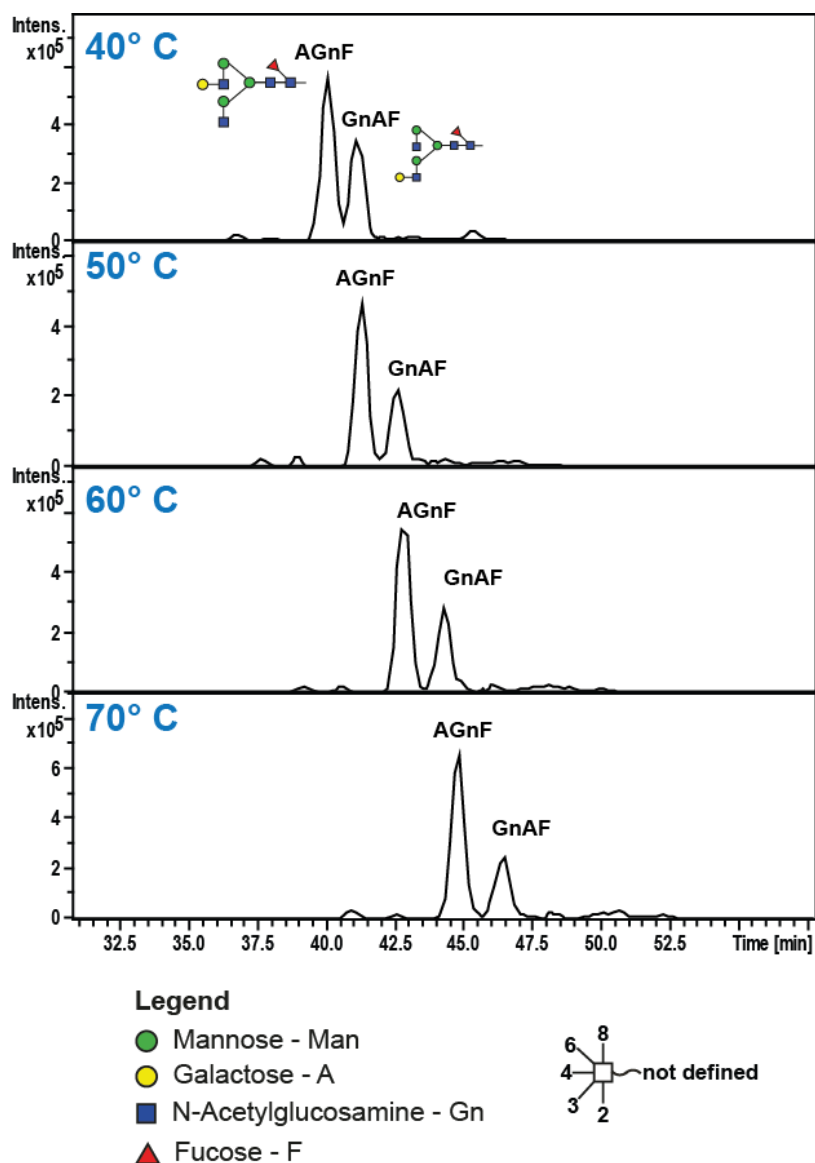


Figure 2.8: Influence of column temperature on isobaric glycan separation by PGC-LC. The retention of N-glycans clearly increased with higher column temperatures, which also resulted in better baseline separation of these two arm isomer N-glycan structures.

In positive mode, applied electrospray voltage is known to cause polarisation of the carbon surface resulting in a total retention of sialylated glycans. This unwanted effect could be avoided by column grounding (18, 19). Therefore, the glycan separation behaviour of grounded and ungrounded columns in negative ionisation PGC-LC ESI MS/MS was evaluated using standard N-glycans derived from bovine fetuin. While the elution pattern of neutral and mono-sialylated glycans were not affected by the grounding status, tri- and tetra-antennary sialylated N-glycans were retained slightly stronger in non-grounded columns. On contrary to positive mode PGC-LC, where a complete retention of these highly

sialylated *N*-glycans was reported. The present negative ion mode PGC-LC ESI MS/MS system used in the analysis did not result in any such effect (Figure 2.9).

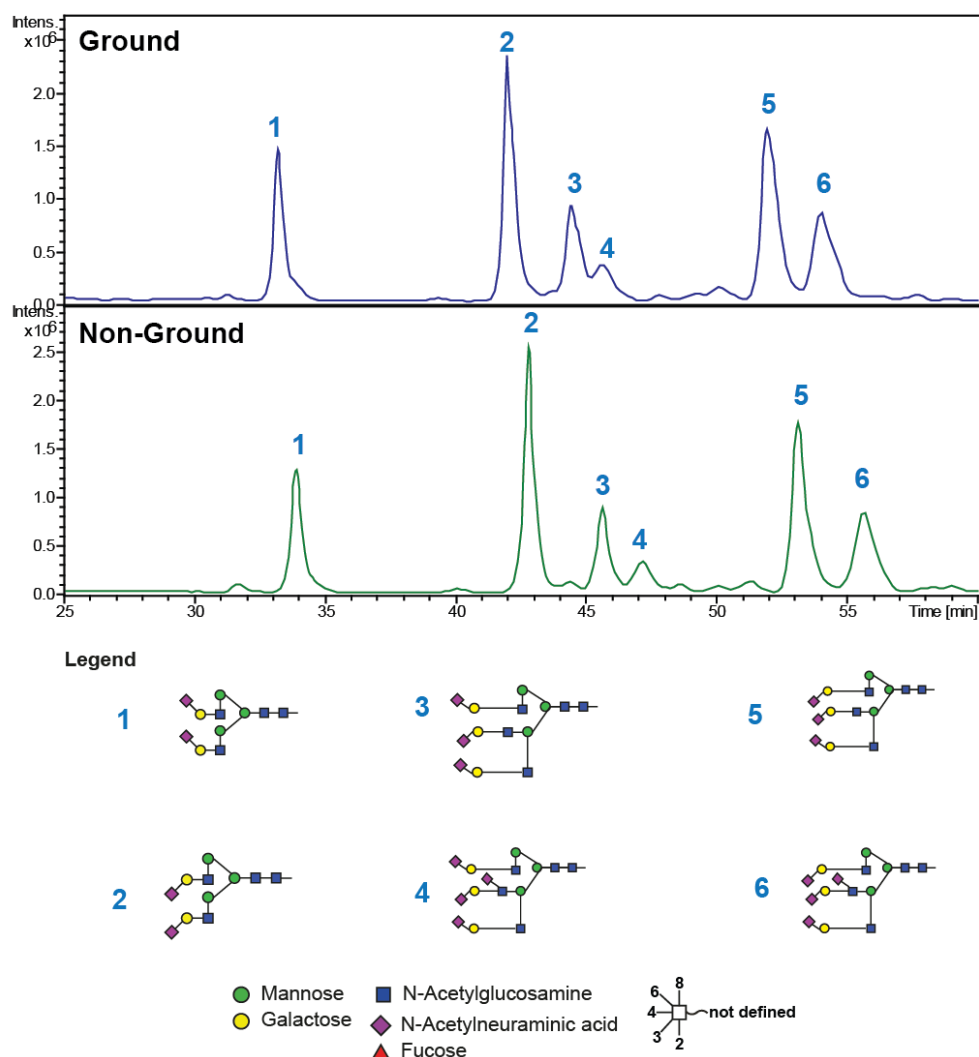


Figure 2.9: Influence of column grounding on *N*-glycan retention in PGC-LC. Sialylated *N*-glycan alditols are retained slightly stronger in non-grounded columns in comparison to grounded columns.

These features (high column temperature and not grounded) could potentially be exploited for improving the analysis of weakly retained oligosaccharides. Very high temperatures, however, should be avoided in glycan analysis to reduce any potential analyte degradation (e.g. sialyl linkages) (15). Despite the absolute LC retention time shifts (either due to eluent or column batch differences); the PGC separation features to separate structural isomers remained highly reproducible (Table 2.5). This unique retention behaviour in combination with individual signature fragment spectra recorded from negatively charged precursor ions was employed to develop a spectra library based matching tool for the automated

identification of glycan structures – a first crucial step in the development of an automated glycan sequencing tool.

Table 2.5: Comparison of absolute and relative retention times of individual example N-glycans determined between 2012-2016. The data clearly showed that relative retention time values are highly in PGC-LC irrespective of unavoidable absolute retention time shifts.

Absolute retention time (min)					Relative retention time			
Year	GnGnF	AGnF	GnAF	AAF	GnGnF	AGnF	GnAF	AAF
2012	37.60	39.30	40.50	42.20	0.89	0.93	0.96	1.00
2013	51.20	54.20	55.10	57.80	0.89	0.94	0.95	1.00
2014	33.70	36.20	37.10	39.20	0.86	0.92	0.94	1.00
2015	28.60	30.70	31.30	33.20	0.86	0.92	0.94	1.00
2016	32.70	35.00	35.80	38.00	0.86	0.92	0.94	1.00
SD	8.68	9.00	9.08	9.37	0.02	0.01	0.01	0.00

2.3.6. BETTER TOGETHER: RELATIVE RETENTION TIME + NEGATIVE ION FRAGMENTATION PROVIDE TWO INDEPENDENT PARAMETERS FOR GLYCAN STRUCTURE ELUCIDATION

The present investigation emphasise that (a) Glycan fragmentation represents one important factor for the development of a spectral matching tool that also requires a similarity scoring between acquired MS/MS fragment and reference spectra (b) glycan fragmentation patterns in negative mode acquired from different ion trap instruments over three independent laboratories is similar (c) glycan isomer separation behaviour by PGC is highly reproducible. Therefore, a global refinement of a spectral matching tool incorporating retention time, associated experimental and MS acquisition metadata within a single tool represents one important step forward to the establishment of a profound tool for automated glycan sequencing.

2.3.6.1. GlycoRRT MS/MS Library

For this purpose a GlycoRRT spectral library was established by incorporating the two distinctive features of negative mode PGC-LC assisted glycomics: a) retention time and b) negative ion fragment spectra. The initial GlycoRRT library developed contained data derived from 40 synthetic *N*-glycan standards (Table 2.6) that included the absolute retention time value for each individual glycan together with the respective negative mode product ion spectra and the observed diagnostic fragment ions, theoretical glycan mass and glycan composition represented in the GlycoMod format. In addition, each entry also contained a glycan name represented in Proglycan nomenclature to distinguish the structural isomers in a single letter code sequence. A representative screen shot entry for the mono galactosylated, core fucosylated *N*-glycan in the GlycoRRT spectral library encompassing all the above mentioned features is shown in (Figure 2.10).

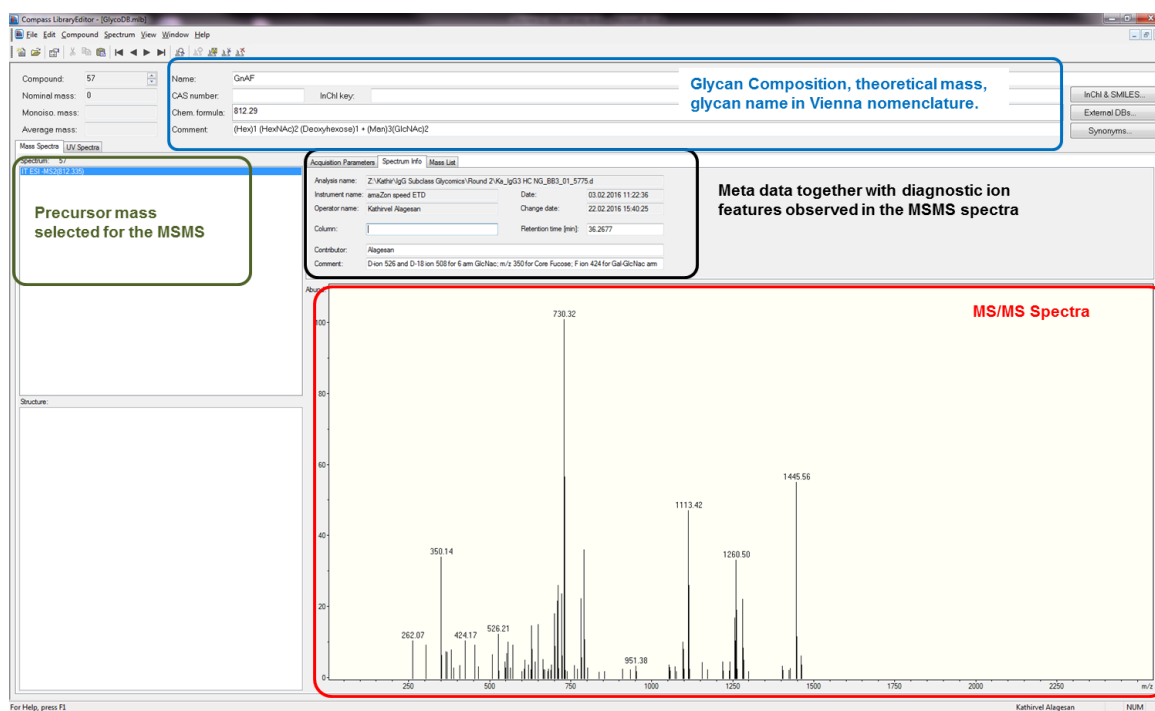


Figure 2.10: GlycoRRT MS library screenshot example for the mono galactosylated fucosylated *N*-glycan (GnAF).

Table 2.6: Overview on the oligosaccharide standards used to develop the GlycoRRT Msn library. Precursor masses ($z=1$ and $z=2$) are shown together with absolute and relative retention time (RRT, using Na(2-6)Na(2-6) [RRT 1] and AAF [RRT 2] N-glycans as reference values).

SNo	Proglycan Name	Precursor mass		Absolute Retention time (min)	Relative retention time (RRT)	
		[M-2H] ²⁻	[M-H] ⁻		RRT1*	RRT2*
1	M1	293.11	587.23	17.50	0.46	0.47
2	M1F	366.14	733.30	27.00	0.71	0.73
3	M2	374.13	749.31	22.30	0.59	0.60
4	M3	455.16	911.37	29.50	0.77	0.80
5	M4	536.19	1073.38	26.90	0.71	0.73
6	M4	536.19	1073.38	32.90	0.86	0.89
7	M5	617.21	1235.57	26.30	0.69	0.71
8	M6	698.24	1397.49	24.60	0.65	0.66
9	M9	941.32	1883.65	23.30	0.61	0.63
10	M9 Glu1	1022.34	2045.69	25.40	0.67	0.68
11	GnGnF	731.27	1463.55	31.90	0.84	0.86
12	GnMan1F	556.74	1114.46	33.40	0.88	0.90
13	Man1GnF	556.74	1114.46	20.90	0.55	0.56
14	Man1Man1F	528.19	1057.39	39.20	1.03	1.06
15	GnGn	658.24	1317.49	24.00	0.63	0.65
16	Gn(GnGn)	759.78	1520.57	29.20	0.77	0.79

17	AGn	739.27	1479.54	26.70	0.70	0.72
18	GnA	739.27	1479.54	27.60	0.72	0.74
19	AA	820.29	1641.60	30.60	0.80	0.82
20	(GnGn)(GnGn)	861.32	1723.63	27.00	0.71	0.73
21	Gn(GnGn)bi	861.32	1723.63	18.70	0.49	0.50
22	GnGnFbi	832.81	1666.63	23.00	0.60	0.62
23	GnGnbi	759.78	1520.57	16.00	0.42	0.43
24	AAF	893.30	1787.65	37.10	0.97	1.00
25	AGnFbi	913.83	1828.68	24.80	0.65	0.67
26	GnAFbi	913.83	1828.68	23.60	0.62	0.64
27	AAFbi	994.86	1990.73	26.40	0.69	0.71
28	A(AA)	1002.86	2006.73	38.20	1.00	1.03
29	(GnGnGn) (GnGn)bi	1064.40	2129.81	14.00	0.37	0.38
30	(AA)(AA)	1185.42	2371.86	39.60	1.04	1.07
31	Na(2-3)A	965.84	1932.69	40.20	1.06	1.08
32	ANa(2-3)	965.84	1932.69	40.60	1.07	1.09
33	Na(2-3) Na(2-3)	1114.00	2223.79	48.70	1.28	1.31
34	A(AF)	893.37	1787.65	35.10	0.92	0.95
35	A(AF)	893.37	1787.65	27.50	0.72	0.74
36	(AF)(AF)	966.35	1933.71	25.40	0.67	0.68

37	Na(2-6) Na(2-6)	1111.40	2223.79	38.10	1.00	1.03
38	ANa(2-6)	965.84	1932.69	33.30	0.87	0.90
39	Na(2-6) A	965.84	1932.69	32.30	0.85	0.87
40	Man3Gnbi	820.34	1438.52	16.60	0.44	0.45
<p>* RRT 1 expressed in relation to Na(2-6)Na(2-6) glycan * RRT 2 expressed in relation to AAF glycan</p>						

2.3.6.2. Glycan RRT Library Evaluation

The feasibility of the spectral matching tool for automated glycan sequencing was first verified by re-searching the dataset used to create the library against the developed GlycoRRT library. The key component of any spectral library searching method is a scoring function that numerically defines the similarity of two spectra (RFit' score). As expected, the re-search of the initial dataset against the library identified the query compound with high confidence as evident by the high RFit' score (Appendix Table 1.1).

After this initial assessment, the GlycanRRT library was further evaluated to identify the *N*-glycans derived from humanised IgG1k expressed in murine suspension culture¹. The obtained protein samples were carbamidomethylated and separated electrophoretically to distinguish the Ig heavy and light chains. Following SDS-PAGE the proteins were transferred onto a PVDF membrane and stained using direct blue 71. Apart from the expected heavy and light chain components an additional band was observed migrating in close proximity to the heavy chain. This additional band could be a result of protein degradation that occurred during sample transport or be derived from co-purified contaminant proteins. Though no protein ID has been performed on the respective band, an in-depth glycan analysis was performed on all labelled major proteins (Figure 2.11).

¹ The Protein samples were provided by Dr. M. Lorna A. De Leoz as a part of NIST Interlaboratory Study on Glycosylation Analysis.

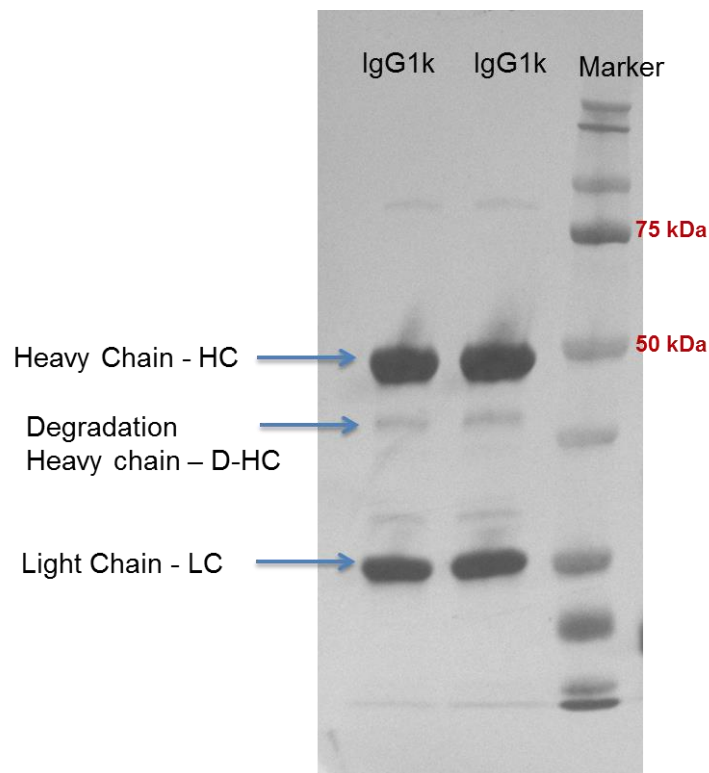


Figure 2.11: Electroblot of 1-D SDS-PAGE separated IgG1k after reduction and alkylation.

The *N*-glycan structures attached to the individual proteins were first identified using the developed GlycoRRT library and subsequently manually validated. In total, 9 different *N*-glycan structures were identified from 500 ng protein using this spectra library approach (Table 2.7). The average RFit' score ranged from 997 for GnGnF to 904 for AGnFbi *N*-glycan indicating high confidence in the structure assignment (Table 2.8).

Any scoring algorithm is in principle prone to false-positive assignment, and spectral matching scoring is no exception. However, such wrong assignments can be significantly reduced when glycan elution profile information is included. The possible variations in absolute retention time differences can easily be avoided using the relative retention time approach. The relative retention time (RRT) for each identified *N*-glycan was manually calculated using the AAF isomer elution time as a reference value of "1". The delta RRT (difference between library RRT and experimentally obtained RRT) identified two potential outliers AGnFbi (-0.36) and GnGnFbi (-0.33) (Table 2.8).

Table 2.7: List of N-glycans identified from recombinant IgG1k expressed in murine suspension culture.

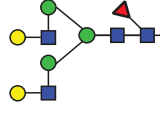
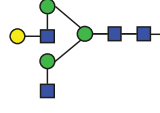
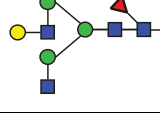
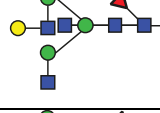
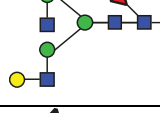

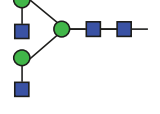
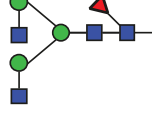
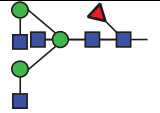
Glycan ID	RT (min)	Precursor m/z	Z	calc. m/z	Δ Mass [Da]	Glycan structure	Proglycan name	HC	D-HC
1	25.4	893.37	2	893.37	0		AAF	X	-
2	18.5	739.33	2	739.27	0.06		AGn	X	-
3	23.2	812.36	2	812.29	0.07		AGnF	X	X
4	26.2	913.91	2	913.83	0.08		AGnFbi	X	-
5	23.9	812.35	2	812.29	0.06		GnAF	X	X
6	15.4	571.34	1	571.23	0.11		GnF	X	-
7	16.7	658.33	2	658.24	0.09		GnGn	X	-
8	21.4	731.33	2	731.27	0.06		GnGnF	X	X
9	24.3	832.87	2	832.81	0.06		GnGnFbi	X	-

Table 2.8: Representative result output exported after glycan identification search (for heavy chain)

#	RT [min]	Compound Name	Spec. Comment	Lib. RT [min]	RFit'	RRT Lib	RRT run	Delta RRT
1	25.4	AAF	D ion 688 and D-18 ion 670; m/z 350 & 571 for Core Fuc; F ion 424 for Gal-GlcNAc, E ion 466 for 3 arm Gal-GlcNAc;	36.8	991	1.00	1.00	0.00
2	18.5	AGn	D ion 688 and D-ion 670 for 6 arm Gal-GlcNAc	26.7	987	0.73	0.73	0.00
3	23.2	AGnF	D-ion 688 and D-18 ion 670 for 6 arm Gal-GlcNAc; m/z 350 for Core Fuc	35.2	988	0.96	0.91	0.04
4	26.2	AGnFbi	D-221 ion 670 for Bisecting GlcNAc and 6 arm Gal-GlcNAc; m/z 350 for Core Fuc	24.7	904	0.67	1.03	-0.36
5	23.9	GnAF	D-ion 526 and D-18 ion 508 for 6 arm GlcNac; m/z 350 for Core Fucose; F ion 424 for Gal-GlcNac arm	36.3	985	0.99	0.94	0.05
6	15.4	GnF	m/z 350 for Core Fuc	21.6	986	0.59	0.61	-0.02
7	16.7	GnGn	D-ion 526 and D-18 ion 508 for 6 arm GlcNAc	25.1	989	0.68	0.66	0.02
8	21.4	GnGnF	D ion 526 and D-18 ion 508 ; m/z 350 for core Fuc; F ion 262 GlcNAc	31.9	997	0.87	0.84	0.02
9	24.3	GnGnFbi	D-221 ion 508 for Bisecting GlcNAc ; m/z 350 for Core Fuc	22.9	918	0.62	0.96	-0.33

False positive assignment obtained in spectra matching can either be due to low-quality library spectra or wrong glycan structure entries that were put into the library. Despite careful quality control, stand-alone spectral matching indicated very high similarity between the query and the reference spectra. The AGnFbi and GnGnFbi *N*-glycan structures were subsequently identified to be GnGnF where the 3-arm antennae was extended carrying additional HexNAc and Hex-HexNAc

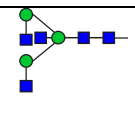
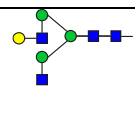
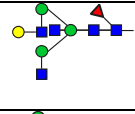
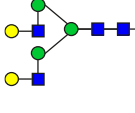
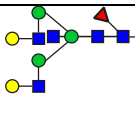
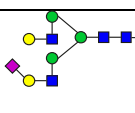
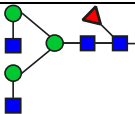
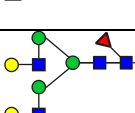
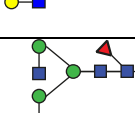
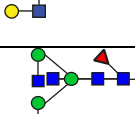
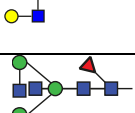
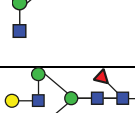
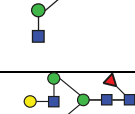
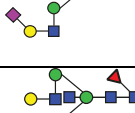
residues, respectively. In the present scenario, the most likely cause for the false positive assignment is derived from the fact that matching reference spectra for these particular structures were initially not present in the library and thus the next closest matches have been selected by the algorithm. Manual verification of the MS/MS data indicated that the very high RFit' scoring is also due to an incorrect dominant isotopic peak matching that resulted in a false positive assignment. One of the main disadvantages of the spectral library searching approach is the fact that only glycans can reliably be identified that have already been annotated and stored in the spectral library. Nevertheless, novel entries can easily be added to the library, and the GlycoRRT approach provides an additional opportunity to identify such initial wrong assignments. In summary, glycan spectral matching and scoring tool (GlycoRRT) including both, negative mode fragmentation and relative retention time represents an important step forward towards reliable automated glycan identification.

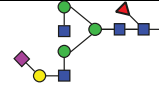
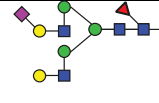
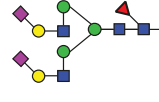
2.3.6.3. LC-MS/MS blind trial – universal applicability of GlycoRRT library

The feasibility of the implementing GlycoRRT in a different laboratory setting was evaluated using the LC-MS data obtained from Prof. Nicolle H. Packer lab at Macquarie University, Sydney². No specific instructions were provided for the data acquisition or the LC gradient. The obtained raw data were processed and the *N*-glycan present in the human IgG was identified using the developed GlycoRRT library and the results are shown in the Table 2.9. In total, 17 different *N*-glycans were identified in less than 3 min demonstrating the universal applicability of the developed GlycoRRT library. The identified glycans were then validated manually and verified indicating 100% accuracy in the automated glycan structural assignment using the developed software.

² Raw LC-MS data provided by Dr. Jodie Abrahams, Macquarie University, Sydney.

Table 2.9: List of N-glycans identified using GlycoRRT library for the LC-MS data obtained from a different laboratory using different LC-MS parameters.

#	RT [min]	Precursor m/z	Compound Name	Glycan structure	Lib. RT [min]	RFit'	Run RRT	Lib RRT	Delta RRT
1	29.9	759.79	GnGnbi		16	676	0.62	0.43	-0.19
2	40	739.24	AGn		26.7	718	0.84	0.73	-0.11
3	37.7	913.83	AGnFbi		24.7	730	0.79	0.67	-0.12
4	43	820.33	AA		30.7	766	0.90	0.83	-0.06
5	39.6	994.87	AAFbi		26	772	0.83	0.71	-0.12
6	45.2	965.87	ANa(2-6)		33.5	789	0.94	0.91	-0.03
7	44.5	731.29	GnGnF		31.9	834	0.93	0.87	-0.06
8	47.9	893.35	AAF		36.8	839	1.00	1.00	0.00
9	46.2	812.32	GnAF		36.3	840	0.96	0.99	0.02
10	38.4	913.86	GnAFbi		23.60	875	0.80	0.64	-0.15
11	36.1	832.83	GnGnFbi		22.9	876	0.75	0.62	-0.13
12	45.4	812.32	AGnF		35.2	880	0.95	0.96	0.01
13	50.1	1038.91	ANaF		35.6	880	1.05	0.97	-0.08
14	42.3	1140.41	ANabi		25.2	883	0.88	0.68	-0.20

15	48	957.86	GnNaF		37.3	908	1.00	1.01	0.01
16	49.4	1038.91	ANaF		35.6	934	1.03	0.97	-0.06
17	50.6	1184.41	Na(2-6) Na(2-6)F		36.7	954	1.06	0.95	-0.06
<i>RRT run and RRT lib expressed in relation to AAF glycan</i>									

2.3.6.4. Annotation of human immunoglobulome glycosylation

The constant region (Fc) of human immunoglobulin (Ig's) mediates most effector functions through isotype and subclass selection while still maintaining antigen specificity. Alterations in Fc glycosylation are known to be associated with the pathogenesis of disease conditions such as autoimmune disease and natural infections (20, 21). These observed changes in Ig glycosylation during disease onset and progression are of immense biological significance as (22, 23) . Availability of high-throughput glycan analysis tools assisting in automated structural characterisation of Ig glycosylation provides one fundamental basis for a better understanding of the functional roles of Fc glycosylation. In addition, detailed structural characterisation of disease associated aberrant glycosylation signatures can help to identify any potential disease biomarkers providing insights into pathogenesis. As a first step towards this end, the developed GlycanRRT library was further expanded using the *N*-glycans released and characterised from human immunoglobulins (Figure 2.12 and 2.13).

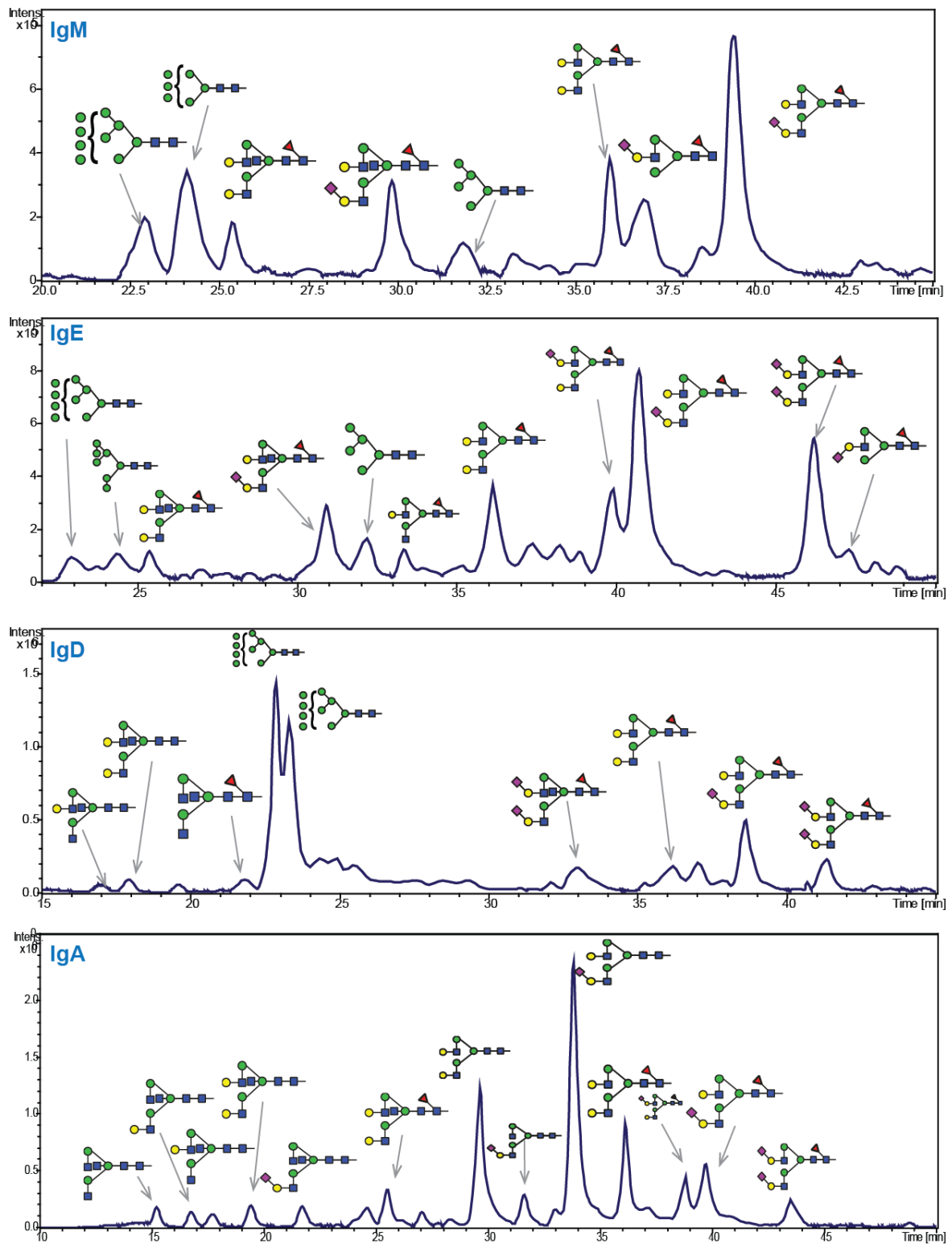


Figure 2.12: N-glycan Base peak chromatogram (BPC) derived from various immunoglobulins isotype (IgM, IgE, IgD and IgA) heavy chains. The major N-glycan structures are annotated in the BPC.

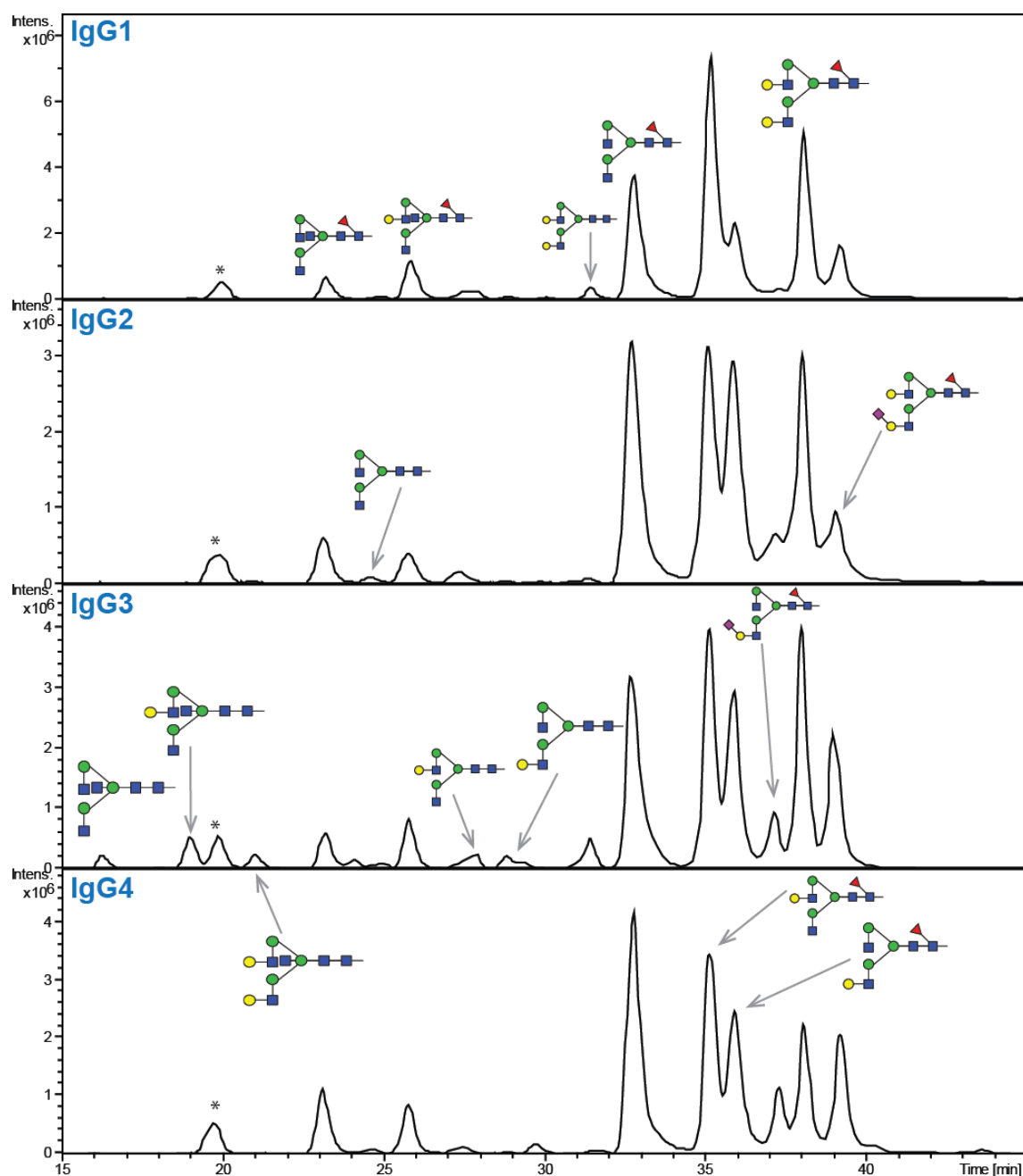


Figure 2.13: N-glycan Base peak chromatogram (BPC) derived from Immunoglobulins G (IgG) subclass heavy chains. The major N-glycan structures are annotated in the BPC. The asterisk indicates a non-glycan contamination signal.

2.3.7. INFLUENCE OF SPS PARAMETER ON QUANTIFICATION

The Smart Parameter Setting (SPS) available in the used ion trap instrument supports in the auto-adjustment of the acquisition mass window to a particular target m/z value. Typically, the SPS target is set in the middle of the intended acquisition mass range. Any influence the SPS target parameter has on glycan identification and quantitation was first evaluated using the Man6 glycan via direct infusion MS analysis. The SPS target was sequentially increased in 100 Da steps

from m/z 900 to 1400 and any in- or decrease in the signal intensities was expressed in relation to the signal intensity observed at SPS 900.

Under the selected MS parameter settings Man6 can be detected either as a singly or doubly charged precursor ion or in both the charge states. The acquired data provided clear evidence that the SPS target m/z also had a direct influence on the detected charge state (Figure 2.14-A & B). This was also explained by the fact that alterations in SPS target did not only optimise the acquisition mass window but also the ion transfer. This became evident by the observed signal intensity decrease of the doubly charged precursor that was compensated by a signal intensity increase of singly charged precursor. The SPS target shift from m/z 900 to 1400 resulted in an approximately 5-fold decrease in the singly charged precursor signal intensity (Figure 2.14-A) whereas the doubly charged precursor signal intensity decreased 0.38-fold (Figure 2.14-B). When both charge states were considered for quantitation, however, the overall relative signal intensity decreased by 0.73-fold when the SPS target m/z was increased from 900 to 1400 (Figure 2.14-C).

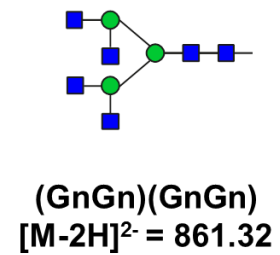
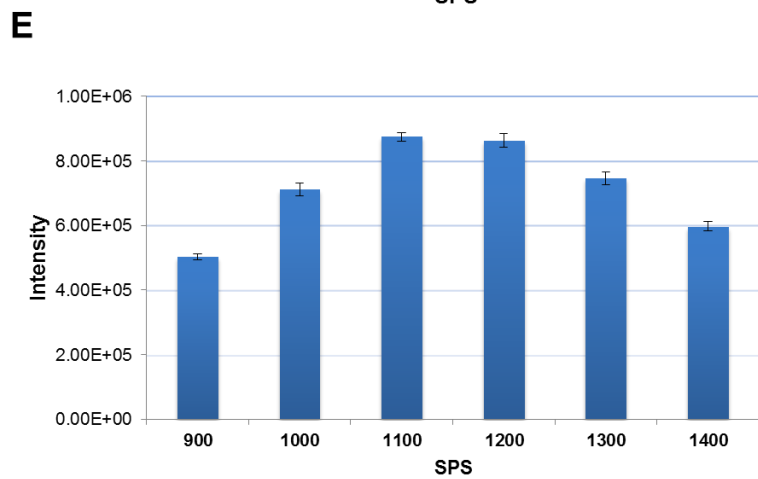
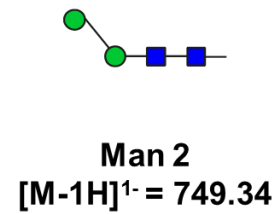
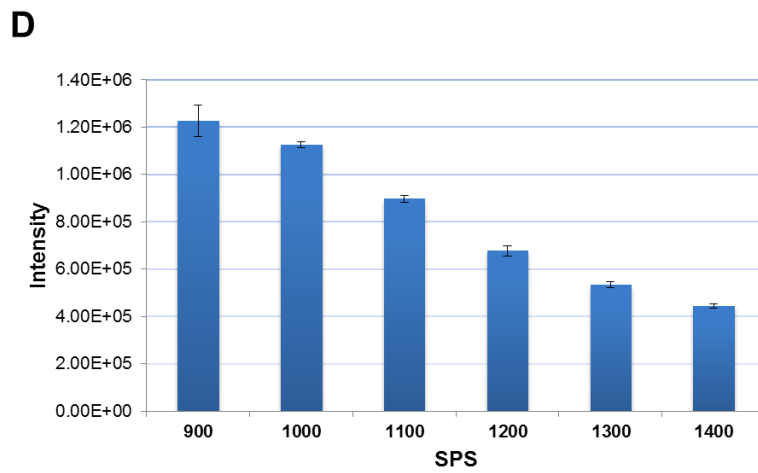
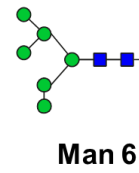
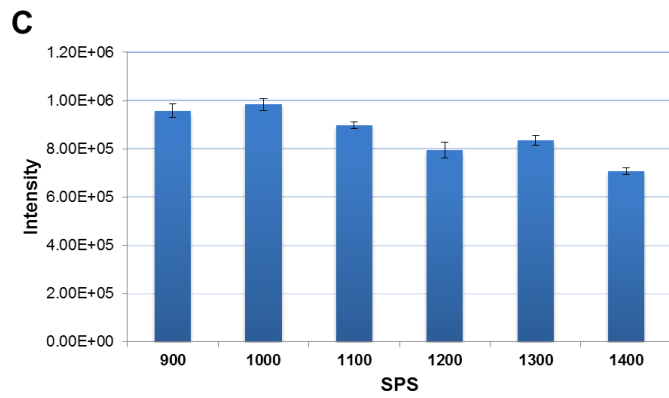
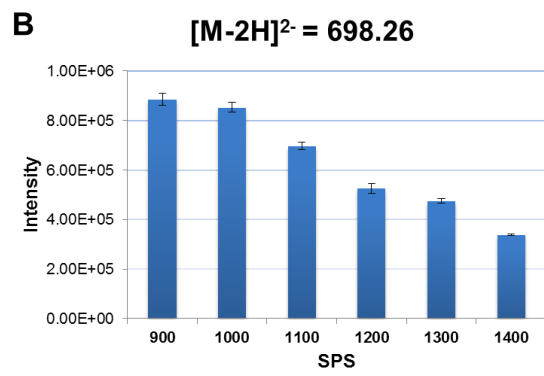
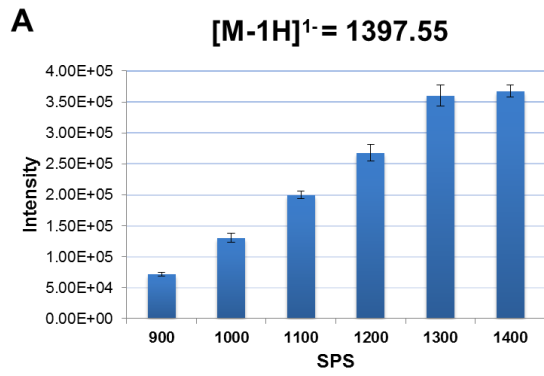
Intrigued by this observation, a panel of available *N*-glycan standards was tested to evaluate the SPS parameter influence on signal intensity and subsequent relative quantitation. In the case of Man2 the shift in SPS target m/z from 900 to 1400 resulted in the gradual decrease in signal intensity (Figure 2.14-D). In the case of tetra-antennary *N*-glycan (GnGn)(GnGn) ($[M-2H]^{2-} = 861.32$), however, the SPS target m/z increase from m/z 900 to 1100 already resulted in 1.67-fold increase in signal intensity. A further increase to 1200 resulted in a 1.73-fold increase. The subsequent increase in SPS target m/z to 1400 resulted in the gradual signal intensity decline that in the end just left a 1.26-fold increase (Figure 2.14-E). A similar trend was observed for the bi-antennary monosialylated *N*-glycan NaA ($[M-2H]^{2-} = 965.84$) (Figure 2.14-F).

An increase in SPS target m/z from 900 to 1400 had a larger effect for the tetra-antennary, bisected *N*-glycan (GnGnGn)(GnGn)bi ($[M-2H]^{2-} = 1064.44$) resulting in an approximate 6.8-fold signal intensity increase (Figure 2.14-G). Considering the observed exceptional SPS parameter dependant influence on signal intensities and detected precursor ion charge it is necessary to develop optimised analysis parameters that depend upon the sample type to avoid any under-representation

of individual glycan structures. The acquired data indicated that the selection of a SPS target m/z oriented towards the higher end of the intended m/z scan range provided better quantitative results compared to the traditional mid-range SPS selection when the glycans over a large m/z range are to be covered.

The use of exoglycosidase enzymes in combination with PGC-LC-ESI-MS has recently been reported to improve the relative quantitation of tri and tetra-antennary *N*-glycan classes (24). However, this approach requires an array of exoglycosidase enzymes and substantial sample amount involving multiple sample preparation steps. Also, information pertaining to individual e.g. NeuAc structure isomers present in the undigested sample is lost. Considering the intrinsic complexity of the glycan, there is no single universal method currently available that will cover each and every aspect in glycan analysis. Thus it is of utmost importance that a method provides suitable versatility to cover a wide m/z and structure range as good as possible. Depending on the target sample, conditions need to be optimised, but once determined for a specific sample set (e.g. body fluid, tissue) these optimised settings can be widely applied.

In that context it is important to note that a comprehensive reporting of all meta-data describing the experimental procedure, instrumentation, instrumental setup, and data acquisition protocols possibly influencing the qualitative and quantitative outcome of glycomics analyses is crucial for appropriate data evaluation. The MIRAGE guidelines that will facilitate the evaluation MS based glycomics experiments are one crucial step towards this goal (25).



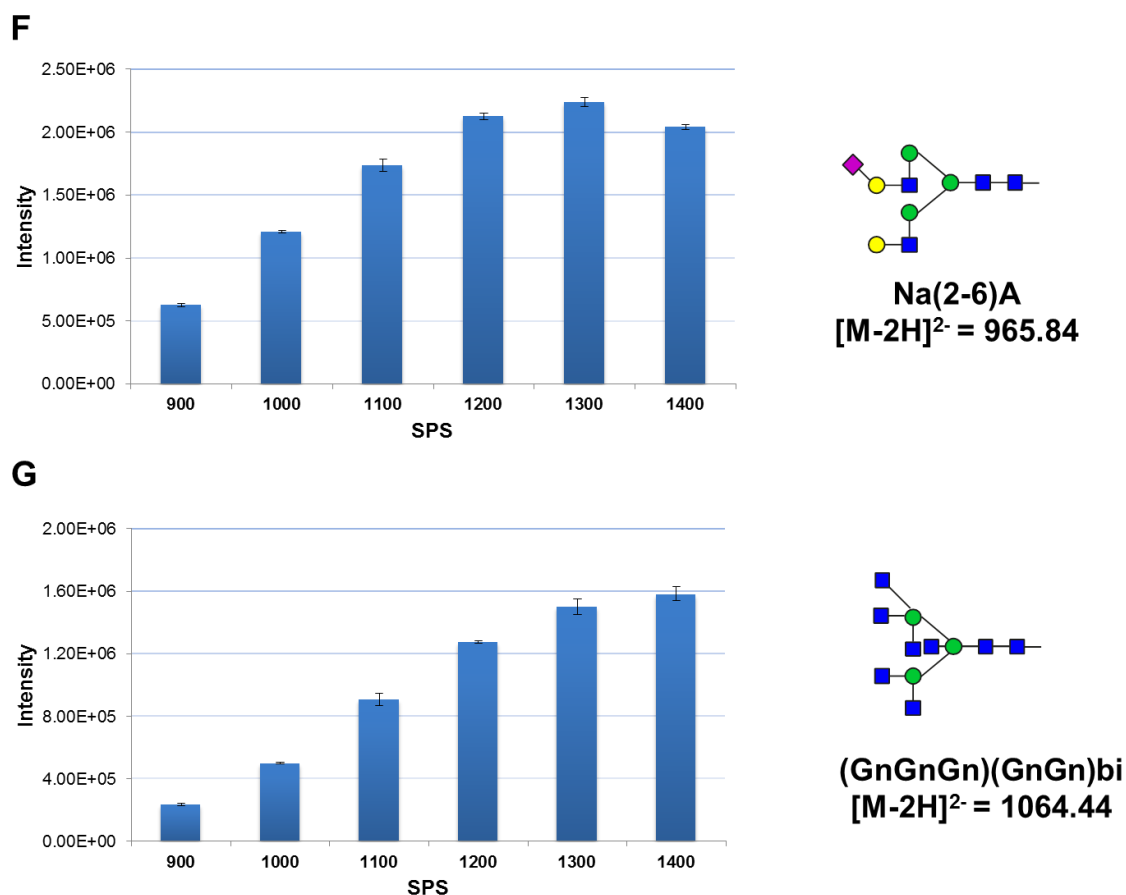


Figure 2.14- A-G: Influence of SPS target m/z on glycan quantitation evaluated by direct infusion MS analysis of reduced N-glycan alditols over the indicated mass ranges. The signal intensities of each N-glycan structure were denoted as the average isotope intensities summed up over 20 sec elution range. Mean and standard deviation of triplicate summed up intensities are shown.

2.3.8. QUANTITATIVE GLOBAL GLYCOMICS

The identified individual N-glycans present in human Immunoglobulins (Ig) and non-depleted human plasma were quantified by integrating the area under the curve (AUC) derived for the respective Extracted ion chromatogram (EIC's). For each measurement, the obtained values were normalised to the sum of all detected N-glycan.

2.3.8.1. Human immunoglobulome glycosylation

As a proof of principle application to test the glycomics method the entire subset of human Immunoglobulin (IgM, IgE, IgD, IgA, IgG 1-4) protein was subjected to a detailed glycomics analysis. As already reported previously each Ig class was found to exhibit distinct glycomic profiles. The identified N-glycans were classified into different classes based on their biosynthetic structure features: oligomannose,

neutral complex or sialylated. Sialylated *N*-glycans was present in all immunoglobulin isotypes and subclasses analysed, though in very different levels. The highest degree of sialylation was detected on IgE (68%) while IgG1 showed the lowest degree of sialylation (10%, Figure 2.15). Neutral complex type *N*-glycans were the major structures present in the all IgG1-4 subclass (>75%). Oligomannose type *N*-glycans were the major structure feature determined for IgD contributing to approximately 47% of the *N*-glycans on IgD. In the case of IgM and IgE 27% and 18.5%, respectively, were contributed by oligomannose *N*-glycans. Where oligomannose type *N*-glycans were not detectable in IgA in the present analysis.

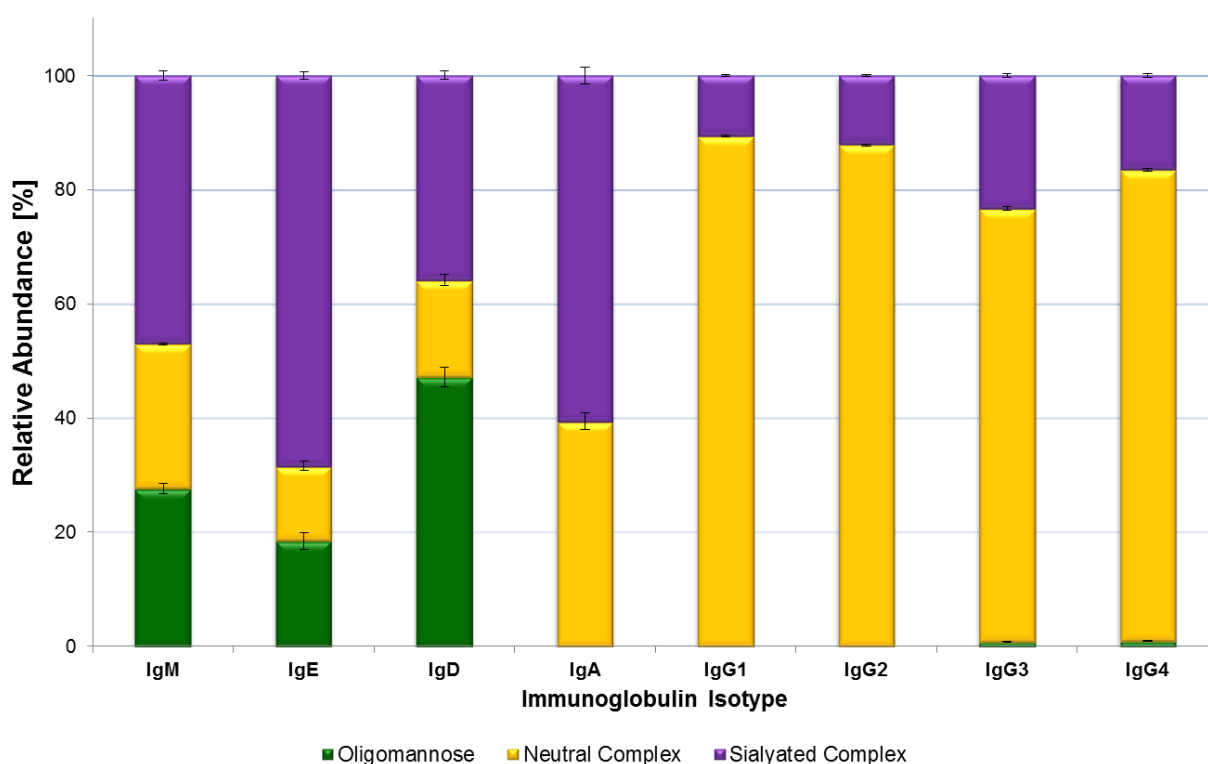


Figure 2.15: Quantitative comparison of different *N*-glycan structure features determined for various Immunoglobulin subtypes (IgM, IgE, IgD, IgA and IgG) using PGC-nLC-ESI-MS/MS analysis.

2.3.8.2. Annotation and quantitation of human plasma *N*-glycome

In total approximately 40 different *N*-glycans were identified from 600 ng protein derived from non-depleted human plasma. The GlycoRRT library approach identified 29 of these 40 *N*-glycans in less than 5 min (Table 2.10) including one potential outlier (Glycan Id # 5). The additional *N*-glycan structures (especially tri-

and tetraantennary *N*-glycans) that were initially not present in the library were identified, manually validated and subsequently added to spectral library, further increasing the number of assigned *N*-glycan structures present in the library (Table 2.11).

Table 2.10: List of *N*-glycans identified from non-depleted human plasma using the automated GlycoRRT search approach.

#	RT [min]	Precursor m/z	Chemical Formula	Compound Name	Glycan (Man)3(GlcNAc)2+	Lib. RT [min]	RFit'	RRT Run	RRT Lib	Delta RRT
1	30.8	1285.98	1285.96	NaNaFbi	(Hex)2 (HexNAc)3 (Deoxyhexose)1 (NeuAc)2	30.4	997	0.82	0.83	0
2	39	1184.43	1184.42	Na(2-6) Na(2-6)F	(Hex)2 (HexNAc)2 (Deoxyhexose)1 (NeuAc)2	36.7	989	1.04	1	0.05
3	33.7	1111.41	1111.4	Na(2-6) Na(2-6)	(Hex)2 (HexNAc)2 (NeuAc)2	34.8	995	0.9	0.95	-0.04
4	31.8	965.85	965.84	Na(2-6)A	(Hex)2 (HexNAc)2 (NeuAc)1	32.6	953	0.85	0.89	-0.04
5	36.8	1184.44	1184.42	Na(2-3) Na(2-3)F	(Hex)2 (HexNAc)2 (Deoxyhexose)1 (NeuAc)2	49.4	992	0.98	1.34	-0.36
6	54.7	1184.42	1184.42	Na(2-3) Na(2-3)F	(Hex)2 (HexNAc)2 (Deoxyhexose)1 (NeuAc)2	49.4	989	1.46	1.34	0.12
7	43	1111.4	1111.4	Na(2-3) Na(2-3)	(Hex)2 (HexNAc)2 (NeuAc)2	40.1	993	1.15	1.09	0.06
8	28.7	783.34	783.27	MNA	(Hex)1 (HexNAc)1 (NeuAc)1	29.5	817	0.77	0.8	-0.03
9	30.4	945.36	945.33	ManNA	(Hex)3 (HexNAc)1 (NeuAc)1	31.2	928	0.81	0.85	-0.03
10	23.9	941.34	941.32	Man9	(Hex)6	23	986	0.64	0.63	0.01
11	23.4	860.33	860.29	Man8	(Hex)5	21.6	965	0.63	0.59	0.04
12	24.7	1397.59	1397.49	Man6	(Hex)3	22.5	986	0.66	0.61	0.05
13	24.8	698.29	698.24	Man6	(Hex)3	24.6	976	0.66	0.67	-0.01
14	26.8	864.36	864.3	M1Na	(Hex)2 (HexNAc)1 (NeuAc)1	27.6	987	0.72	0.75	-0.03
15	36.5	957.87	957.84	GnNaF	(Hex)1 (HexNAc)2 (Deoxyhexose)1 (NeuAc)1	37.3	980	0.98	1.01	-0.04
16	30.4	884.85	884.81	GnNA	(Hex)1 (HexNAc)2 (NeuAc)1	31.1	993	0.81	0.85	-0.03
17	22.7	832.87	832.81	GnGnFbi	(HexNAc)3 (Deoxyhexose)1	22.9	994	0.61	0.62	-0.02

18	32.1	731.32	731.27	GnGnF	(HexNAc)2 (Deoxyhexose)1	31.9	991	0.86	0.87	-0.01
19	32.2	1463.6	1463.55	GnGnF	(HexNAc)2 (Deoxyhexose)1	32.5	984	0.86	0.88	-0.02
20	35.4	812.34	812.29	GnAF	(Hex)1 (HexNAc)2 (Deoxyhexose)1	36.3	990	0.95	0.99	-0.04
21	38.7	1038.87	1038.87	ANa(2-6)F	(Hex)2 (HexNAc)2 (Deoxyhexose)1 (NeuAc)1	35.6	970	1.03	0.97	0.07
22	47.4	1038.88	1038.87	ANa(2-3)F	(Hex)2 (HexNAc)2 (Deoxyhexose)1 (NeuAc)1	35.6	907	1.27	0.97	0.3
23	23.6	1067.38	1067.38	ANabi	(Hex)2 (HexNAc)3 (NeuAc)1	24.1	972	0.63	0.65	-0.02
24	28.9	1140.42	1140.41	ANabiF	(Hex)2 (HexNAc)3 (Deoxyhexose)1 (NeuAc)1	26.3	965	0.77	0.71	0.06
25	32.6	965.87	965.84	ANa(2-6)	(Hex)2 (HexNAc)2 (NeuAc)1	33.5	925	0.87	0.91	-0.04
26	25.3	913.88	913.83	AGnFbi	(Hex)1 (HexNAc)3 (Deoxyhexose)1	24.7	992	0.68	0.67	0.01
27	34.7	812.34	812.29	AGnF	(Hex)1 (HexNAc)2 (Deoxyhexose)1	35.2	980	0.93	0.96	-0.03
28	26.7	994.86	994.86	AAFbi	(Hex)2 (HexNAc)3 (Deoxyhexose)1	26	904	0.71	0.71	0.01
29	37.4	893.36	893.37	AAF	(Hex)2 (HexNAc)2 (Deoxyhexose)1	36.8	983	1	1	0
<i>RRT run and RRT lib expressed in relation to AAF glycan.</i>										

Table 2.11: List of additional N-glycans identified from non-depleted human plasma by manual annotation.

#	RT [min]	Precursor m/z	Chemical Formula	Compound Name	Cmpd. Comment
30	51.8	1178.64	1178.06	4XNa Core	(Hex)4 (HexNAc)4 (NeuAc)4 + (Man)3(GlcNAc)2
31	41.9	1439.51	1439.5	3XNa Core	(Hex)3 (HexNAc)3 (NeuAc)3 + (Man)3(GlcNAc)2
31- a	42.1	959.34	1439.5	3XNa Core	(Hex)3 (HexNAc)3 (NeuAc)3 + (Man)3(GlcNAc)2
32	44.5	1439.5	1439.5	3XNa Core	(Hex)3 (HexNAc)3 (NeuAc)3 + (Man)3(GlcNAc)2
32- a	44.9	959.33	1439.5	3XNa Core	(Hex)3 (HexNAc)3 (NeuAc)3 + (Man)3(GlcNAc)2
33	52.5	1439.48	1439.5	3XNa Core	(Hex)3 (HexNAc)3 (NeuAc)3 + (Man)3(GlcNAc)2
34	39.9	1512.53	1512.57	3XNa CoreFucose	(Hex)3 (HexNAc)3 (Deoxyhexose)1 (NeuAc)3 + (Man)3(GlcNAc)2
34- a	48.5	1512.4	1512.57	3XNa CoreFucose	(Hex)3 (HexNAc)3 (Deoxyhexose)1 (NeuAc)3 + (Man)3(GlcNAc)2
35	35.9	1293.95	1293.95	(2XNa)(AAA)	(Hex)3 (HexNAc)3 (NeuAc)2 + (Man)3(GlcNAc)2

36	40	1293.96	1293.95	(2XNa)(AAA)	(Hex)3 (HexNAc)3 (NeuAc)2 + (Man)3(GlcNAc)2
37	40.9	1293.96	1293.95	(2XNa)(AAA)	(Hex)3 (HexNAc)3 (NeuAc)2 + (Man)3(GlcNAc)2
38	44.2	1293.95	1293.95	(2XNa)(AAA)	(Hex)3 (HexNAc)3 (NeuAc)2 + (Man)3(GlcNAc)2
39	46.1	1293.93	1293.95	(2XNa)(AAA)	(Hex)3 (HexNAc)3 (NeuAc)2 + (Man)3(GlcNAc)2
40	47.8	1184.41	1184.42	Na(2-6) Na(2-3)F MIX	(Hex)2 (HexNAc)2 (Deoxyhexose)1 (NeuAc)2+ (Man)3(GlcNAc)2

The identified *N*-glycans were classified based on the same biosynthetic relevant structure features as it was done for the Immunoglobulins. In agreement with numerous previous reports (26-29), the most abundant *N*-glycan was the NaNa glycan carrying α 2,6 linked Neu5Ac contributing to approximately 54% of the total plasma *N*-glycan pool, which is, however, in contrast to the work published by Song *et al.* who reported the major structure to be a bi-antennary, doubly α 2,3 linked Neu5Ac carrying *N*-glycan. In this work the presumed α 2,3 was determined using α 2-3 neuraminidase digests (30). Nevertheless, despite its high specificity for α 2-3 linked sialic acids, α 2-3 Neuraminidase is also known to catalyse the hydrolysis of α 2-6 linked sialic acids, though at a slower rate³. Thus it is conceivable that either prolonged incubated times or an excess of enzyme in the incubation mixture could have led to the digestion of the actual α 2,6 linked Neu5Ac residues, subsequently resulting in a wrong structural assignment. In the here present study the relative retention time and distinctive negative mode fragmentation clearly indicated that the most abundant *N*-glycan structure in the human plasma was the doubly α 2,6 sialylated, di-antennary *N*-glycan. The quantitative analysis indicated that sialylated *N*-glycans accounted to approximately 90% of the total *N*-glycome in human plasma (Figure 2.16-A). The sialic acid containing *N*-glycans were further classified into different sub-class based upon the number of sialic acid residues. The majority of sialylated *N*-glycans were di-sialylated (75%), followed by mono- and tri-sialylated *N*-glycans (Figure 2.16-B). Tetra-sialylated *N*-glycans were just present in trace amounts.

³ <https://www.neb.com/products/p0728-2-3-neuraminidase> accessed on 07.11.2016

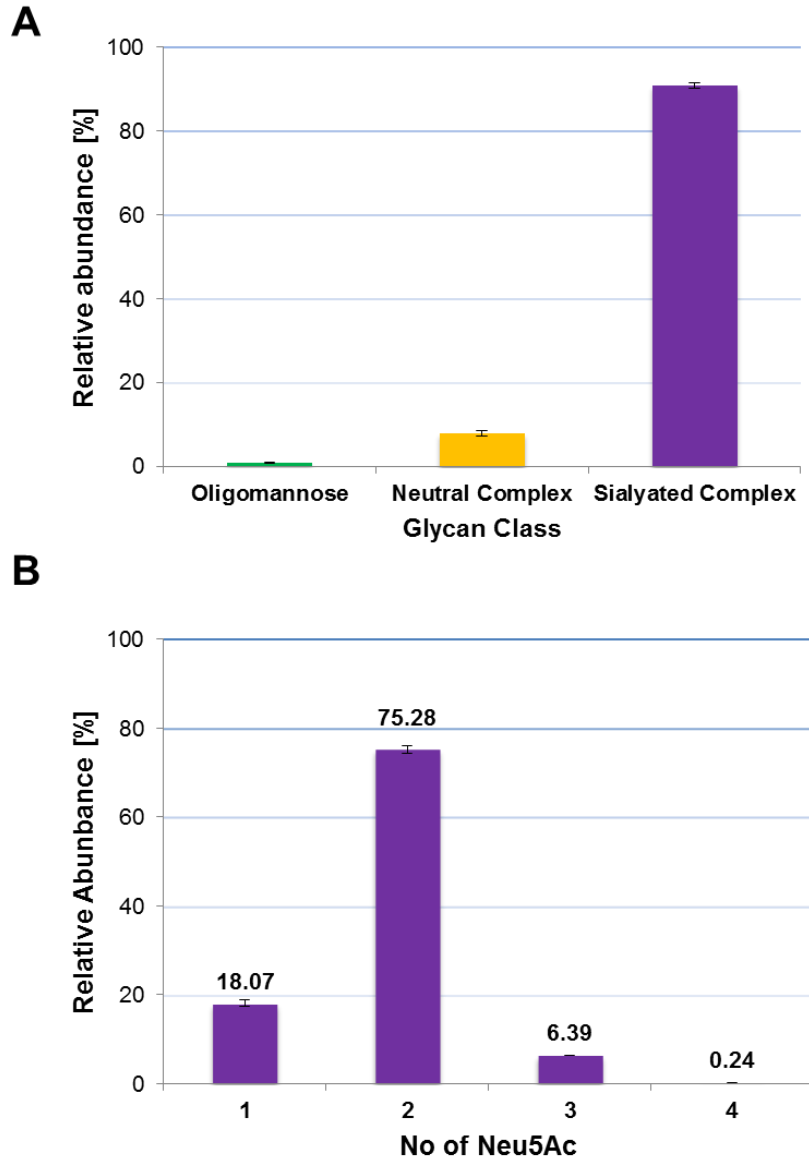


Figure 2.16: (A) Overall relative abundances of various N-glycan structure features identified in human plasma using PGC-nLC-ESI-MS/MS analysis (B) Quantitative comparison of sialylated complex N-glycans based upon the number of Neu5Ac residues.

2.4. CONCLUSION

A spectral library based approach for automated glycan identification in combination with relative retention time has been developed for PGC-LC-ESI-MS/MS based glycomics. All N-glycan structures present in the library were characterised using (i) molecular monoisotopic mass and (ii) CID-MS/MS *de novo* sequencing. Overall the library now contains more than 200 annotated N-glycans structures derived from synthetic N-glycan standards and human Immunoglobulins. The described tool is fully customisable, easily adaptable and

can be readily employed for automated *N*-glycan identification and characterisation. Evaluation of both in-house data and a blind data set obtained from a different laboratory indicated the universal applicability of the developed GlycoRRT library. Development of a similar tool for automated *O*-glycan analysis is in progress.

2.5. REFERENCE

1. Kolarich D, Packer NH. Mass Spectrometry for Glycomics Analysis of N- and O-Linked Glycoproteins. In: Yuriev E, Ramsland PA, editors. *Structural Glycobiology*. Boca Raton, FL 33487-2742: CRC Press; 2012. p. 141-61.
2. Kolarich D, Lepenies B, Seeberger PH. Glycomics, glycoproteomics and the immune system. *Curr Opin Chem Biol*. 2012;16(1-2):214-20.
3. Robinson S, Routledge A, Thomas-Oates J. Characterisation and proposed origin of mass spectrometric ions observed 30 Th above the ionised molecules of per-O-methylated carbohydrates. *Rapid Commun Mass Spectrom*. 2005;19(24):3681-8.
4. Campbell MP, Nguyen-Khuong T, Hayes CA, Flowers SA, Alagesan K, Kolarich D, 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):108-16.
5. Hayes CA, Karlsson NG, Struwe WB, Lisacek F, Rudd PM, Packer NH, et al. UniCarb-DB: a database resource for glycomic discovery. *Bioinformatics*. 2011;27(9):1343-4.
6. Jensen PH, Karlsson NG, Kolarich D, Packer NH. Structural analysis of N- and O-glycans released from glycoproteins. *Nat Protoc*. 2012;7(7):1299-310.
7. Kolarich D, Windwarder M, Alagesan K, Altmann F. Isomer-Specific Analysis of Released N-Glycans by LC-ESI MS/MS with Porous Graphitized Carbon. *Methods Mol Biol*. 2015;1321:427-35.
8. Harvey DJ. Fragmentation of negative ions from carbohydrates: part 3. Fragmentation of hybrid and complex N-linked glycans. *J Am Soc Mass Spectrom*. 2005;16(5):647-59.
9. Everest-Dass AV, Abrahams JL, Kolarich D, Packer NH, Campbell MP. Structural feature ions for distinguishing N- and O-linked glycan isomers by LC-ESI-IT MS/MS. *J Am Soc Mass Spectrom*. 2013;24(6):895-906.
10. Everest-Dass AV, Kolarich D, Campbell MP, Packer NH. Tandem mass spectra of glycan substructures enable the multistage mass spectrometric identification of determinants on oligosaccharides. *Rapid Commun Mass Spectrom*. 2013;27(9):931-9.
11. Harvey DJ, Royle L, Radcliffe CM, Rudd PM, Dwek RA. Structural and quantitative analysis of N-linked glycans by matrix-assisted laser desorption ionization and negative ion nanospray mass spectrometry. *Anal Biochem*. 2008;376(1):44-60.
12. Pabst M, Bondili JS, Stadlmann J, Mach L, Altmann F. Mass + retention time = structure: a strategy for the analysis of N-glycans by carbon LC-ESI-MS and its application to fibrin N-glycans. *Anal Chem*. 2007;79(13):5051-7.
13. Stein SE, Scott DR. Optimization and testing of mass spectral library search algorithms for compound identification. *J Am Soc Spectrom*. 1994;6(9):1044-0305.
14. Campbell MP, Nguyen-Khuong T, Hayes CA, Flowers SA, Alagesan K, Kolarich D, et al. Validation of the curation pipeline of UniCarb-DB: Building a global glycan reference MS/MS repository. *Bba-Proteins Proteom*. 2014;1844(1):108-16.
15. Palmisano G, Larsen MR, Packer NH, Thaysen-Andersen M. Structural analysis of glycoprotein sialylation -part II: LC-MS based detection. *Rsc Adv*. 2013;3(45):22706-26.

16. West C, Elfakir C, Lafosse M. Porous graphitic carbon: A versatile stationary phase for liquid chromatography. *J Chromatogr A*. 2010;1217(19):3201-16.
17. Melmer M, Stangler T, Premstaller A, Lindner W. Comparison of hydrophilic-interaction, reversed-phase and porous graphitic carbon chromatography for glycan analysis. *J Chromatogr A*. 2011;1218(1):118-23.
18. Pabst M, Altmann F. Influence of electrosorption, solvent, temperature, and ion polarity on the performance of LC-ESI-MS using graphitic carbon for acidic oligosaccharides. *Anal Chem*. 2008;80(19):7534-42.
19. Tornkvist A, Nilsson S, Amirkhani A, Nyholm LM, Nyholm L. Interference of the electrospray voltage on chromatographic separations using porous graphitic carbon columns. *Journal of Mass Spectrometry*. 2004;39(2):216-22.
20. Varki A, Freeze HH. Glycans in Acquired Human Diseases. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al., editors. *Essentials of Glycobiology*. 2nd ed. Cold Spring Harbor (NY)2009.
21. Epp A, Sullivan KC, Herr AB, Strait RT. Immunoglobulin Glycosylation Effects in Allergy and Immunity. *Curr Allergy Asthma Rep*. 2016;16(11):79.
22. Raju TS. Terminal sugars of Fc glycans influence antibody effector functions of IgGs. *Curr Opin Immunol*. 2008;20(4):471-8.
23. Jiang XR, Song A, Bergelson S, Arroll T, Parekh B, May K, et al. Advances in the assessment and control of the effector functions of therapeutic antibodies. *Nat Rev Drug Discov*. 2011;10(2):101-11.
24. Abrahams JL, Packer NH, Campbell MP. Relative quantitation of multi-antennary N-glycan classes: combining PGC-LC-ESI-MS with exoglycosidase digestion. *The Analyst*. 2015;140(16):5444-9.
25. Kolarich D, Rapp E, Struwe WB, Haslam SM, Zaia J, McBride R, 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):991-5.
26. Hennig R, Cajic S, Borowiak M, Hoffmann M, Kottler R, Reichl U, et al. Towards personalized diagnostics via longitudinal study of the human plasma N-glycome. *Biochim Biophys Acta*. 2016;1860(8):1728-38.
27. Xia B, Zhang W, Li X, Jiang R, Harper T, Liu R, et al. Serum N-glycan and O-glycan analysis by mass spectrometry for diagnosis of congenital disorders of glycosylation. *Anal Biochem*. 2013;442(2):178-85.
28. Clerc F, Reiding KR, Jansen BC, Kammeijer GS, Bondt A, Wuhrer M. Human plasma protein N-glycosylation. *Glycoconj J*. 2016;33(3):309-43.
29. Jansen BC, Bondt A, Reiding KR, Lonardi E, de Jong CJ, Falck D, et al. Pregnancy-associated serum N-glycome changes studied by high-throughput MALDI-TOF-MS. *Sci Rep*. 2016;6:23296.
30. Song T, Aldredge D, Lebrilla CB. A Method for In-Depth Structural Annotation of Human Serum Glycans That Yields Biological Variations. *Anal Chem*. 2015;87(15):7754-62.

3. DEVELOPMENT OF A NOVEL, ULTRASENSITIVE APPROACH FOR QUANTITATIVE CARBOHYDRATE COMPOSITION AND LINKAGE ANALYSIS

3.1 INTRODUCTION

The aim of functional glycomics is to understand how glycosylation influences the functions of individual glycoproteins. Thus in-depth knowledge is required not just on glycan structures but also their constitutive monosaccharide identities and linkages, which by mass spectrometry based methods are difficult if not impossible to differentiate when entire glycans are analysed. In a comprehensive glycan structure analysis various aspects need to be considered: i) monosaccharide composition, ii) oligosaccharide sequence, iii) type of sugar linkage, iv) branching sites positions and v) elucidation of anomeric configurations need to be realised, usually from a minimum amount of material (1, 2). As already described for the PGC-glycomics approach in *chapter 2* various methodologies are available that provide different aspects of this information either directly on intact glycan structures or after hydrolysis of these into their monosaccharide building blocks. What type of information can be obtained also depends on factors such as pre-existing knowledge on the biosynthetic machinery, the selected analytical approach(es) and last but not least, the available amount and purity of the target molecules. In eukaryotic model species and mammals in particular these pathways are comparably well understood. Studying glycosylation in less well known species, however, very often just provides selected information on glycan composition and structure, as many monosaccharide building blocks are indistinguishable by mass alone and detailed linkage information is also not easily obtained by purely MS/MS based analyses from minute amounts of material (see *also chapter 2*). Monosaccharide composition and linkage analyses of the individual monosaccharide constituents are, nonetheless, required for in-depth structure determination in particular when less well studied organisms such as bacteria or evolutionary more distant animals outside the Mammalia genus are of interest.

The use of specific exoglycosidases represents one possibility to elucidate monosaccharide identity and linkage. Monitoring their activity by different analytical tools such as mass spectrometry, HPLC or Capillary Gel Electrophoresis coupled with Laser Induced Fluorescence detection (CE-LIF) does provide information on oligosaccharide sequence, monosaccharide identity, anomericity and linkage as these enzymes are highly substrate specific (3, 4). However, this method requires a vast array of enzymes, sufficient amounts of material and is time consuming to achieve complete sequencing. In addition, this approach fails when suitable exoglycosidases are unavailable.

Numerous analytical methods have been developed to identify and quantitate monosaccharides from glycoconjugates. Several decades ago gas chromatography interfaced with electron impact ionization mass spectrometry (GC-MS) has been established for monosaccharide determination and is still considered a state of the art method for this purpose. This method is based on the conversion of monosaccharides into partially methylated alditol acetals (PMAAs) that are obtained after a series of derivatisation steps: permethylation, acid hydrolysis and reduction followed by acetylation of partially methylated sugar alditols. Different types of monosaccharide residues and their linkages can be identified based upon the GC retention time and their characteristic fragmentation patterns (5-7).

In addition, alternative methods for monosaccharide analysis are also available. With the exception of HPAE-PAD (8) they usually include some type of chemical derivatisation on the reducing end to improve monosaccharide HPLC separation, detection and analysis. Several methods have in addition been reported for the LC-MS analyses of monosaccharides via post column addition of Na^+ , Cs^+ and NH_4^+ to facilitate the generation of positively charged molecules and I^- and Cl^- when the compounds of interest are to be detected as negatively charged ions (9-14). Also, few LC-MS based approaches have been reported for underivatized monosaccharide compounds (15-17), but many of these approaches lack sensitivity or do not provide valuable linkage information. A major drawback of HPAE-PAD technique is the fact that a dedicated instrument is required. Surprisingly, to date only little effort has been made to develop LC-MS based, highly sensitive and selective methods for unambiguous identification of

monosaccharide composition and their linkages. Here, a novel, highly sensitive nanoLC-ESI ion trap tandem mass spectrometry based method for monosaccharide composition and linkage analysis is presented.

3.2 MATERIALS AND METHODS

3.2.1. Reagents

- L(-)Fucose (Sigma-Aldrich, cat. no. F-2252)
- D(+)Xylose (M)
- D(+)Mannose (Sigma-Aldrich, cat. no. M-4625)
- D(+)Galactose (Sigma-Aldrich, cat. no. G-0750)
- D(+)Glucose (Sigma-Aldrich, cat. no. G-8270)
- N-Acetyl-D-glucosamine (Sigma-Aldrich, cat. no. A8625)
- N-Acetyl-D-galactosamine (Sigma-Aldrich, cat. no. A2795)
- N-Acetyl-D-mannosamine (Sigma-Aldrich, cat. no. A8176)
- N-Acetylneuraminic acid (ACROS Organics cat. no. 227040250)
- DMSO (ACROS Organics, cat. no. 127790010)
- CH₃I (Sigma-Aldrich, cat. no. 03810)
- NaOH (Sigma-Aldrich, cat. no. S5881)
- Glacial acetic acid (Merck, cat. no. 100066)
- TFA (Merk, 1081780050)
- Sodium cyanoborohydride (Sigma-Aldrich, cat. no. 156159)
- DCM (Merck, cat. no. 6048)
- Chloroform (Merck, cat. no. 102445)
- N-glycan standards (Dextra Reading, UK)
- ZipTips (Millipore)

Water was used after purification with a Milli Q-8 direct system (Merck 119 KGaA, Darmstadt, Germany). Stock solutions with a concentration of 1 mM of each monosaccharide (Glc, Gal, Man, GlcNAc, ManNAc, GalNAc, Xyl, Fuc, Neu5Ac) were prepared using MilliQ-8 water.

3.2.2. General procedure for carbohydrate derivatisation

Monosaccharide permethylation was performed according to the procedure described previously by Ciucanu & Kerek (18) with minor modifications as

described by Ciucanu I & Costello CE (19). Briefly, approx. 5 mg of the glycan sample (standard monosaccharides) was dissolved in 1000 μL of DMSO by gentle vortexing, to which 50 mg of finely powdered NaOH was added. The mixture was sonicated for 10 min at room temperature with further incubation of 30 min with occasional shaking. Subsequently, 100 μL of methyl iodide was added to the sample with sonication for another 10 min. Reaction was terminated by addition of 1000 μL water. Permethylated monosaccharides were extracted using 1000 μL of chloroform. The chloroform layer was further washed with equal volumes of water for at least 3 times to remove any residual salts. The organic phase containing the permethylated monosaccharides was dried in vacuo and reconstituted in 20% (v/v) aqueous acetonitrile. Standard monosaccharides were labelled by reductive amination using 2-aminobenzamide (2-AB) as described earlier (20). Excess reagents were removed as described by Pabst et al. previously (21). Briefly, labelling reagent was prepared by dissolving 0.35 M of 2-AB and 1 M sodium cyanoborohydride in 30% acetic acid in DMSO. About 5 mg of monosaccharides were dissolved in 100 μL of the labelling reagent and the reaction mixture was incubated at 65°C for 120 min.

A detailed step-wise developed procedure for the monosacchride linkage and compositional analysis is provided in the section 3.4.

3.2.3. ESI-MS/MS and LC-ESI-MS/MS analysis of the derivatised monosaccharides

All mass spectrometric analyses were performed on an amaZon ETD Speed ion trap (Bruker, Bremen, Germany) coupled to an Ultimate 3000 UHPLC system (Dionex, Part of Thermo Fisher Scientific, Germany).

Differently derivatised monosaccharide standards was dissolved in 30% acetonitrile containing 0.1% formic acid (FA) at the calculated concentration of 50 picomoles/ μL . All the samples were injected for analysis by ESI-MS using direct infusion with a syringe pump at a flow rate of 1 $\mu\text{L}/\text{min}$. The instrument was set up to perform CID fragmentation in positive mode on the selected precursors. Precursor ions were selected manually with the isolation width of $\pm 1\text{Da}$ and fragmentation energy was increased manually to determine the optimal fragmentation energies. The data was recorded in the instrument's "enhanced

mode resolution". Specific instrumental operational parameters used in the present investigation are listed in Table 3.1.

Table 3.1: MS operational parameters used in the present investigation

<i>Ionization mode</i>	<i>Positive mode ESI for monosaccharides</i>
<i>Capillary exit</i>	1-1.3 kV
<i>ICC</i>	On
<i>Maximum accumulation time</i>	50 ms
<i>Target</i>	200,000 (MS/MS)
<i>Scan range</i>	50-500 m/z (for MS and MS/MS)
<i>Isolation window</i>	2.0 m/z
<i>MS/MS fragmentation amplitude</i>	60.0 %
<i>Smart fragmentation option</i>	On (start amplitude 30%—end amplitude 200%)

On the nanoLC system the samples were injected (analyte concentration: 50 pmol) onto the precolumn in 100% solvent A (0.1% formic acid) at the flow rate of 6 μ L/min and unbound components were eluted for 5 min in buffer A. Meanwhile the analytical column was equilibrated in 2% solvent B (acetonitrile with 0.1% formic acid) and a linear gradient was applied using an increasing solvent B concentration at the flow rate of 300 nL/min as follows: steep increase of buffer B from 1 to 13% (from 5 min to 8 min), followed by a slow increase of buffer B from 13% to 30% (8 to 70 min before a steep increase to 90% (70 to 75 min) was executed. The column was held at 90% B for 5 min before re-equilibrating the analytical column in 1% solvent B. Meanwhile the precolumn was re-equilibrated in 100% solvent A before injection of the next sample.

3.3 RESULTS AND DISCUSSION

3.3.1. Rationale and Method development

The aim was to develop an easily adaptable and highly sensitive LC-ESI MS/MS based monosaccharide analysis approach that also provided composition and linkage information on glycoconjugates. To achieve the highest possible sensitivity first the various derivatisation and labelling steps were optimised using standard monosaccharides (Figure 3.1). As part of this method development and optimisation permethylated monosaccharides were analysed after being labelled with fluorescent tags on the reducing end (Table 3.2). Next, the fragmentation behaviour and chromatographic properties of the derivatised monosaccharides were evaluated using various stationary phases (reverse phase C18, porous graphitized carbon, PGC) and the chromatographic conditions were optimised to separate the various monosaccharide stereoisomers.

Table 3.2: Protonated $[M+H]^+$ mass values of the differently derivatised monosaccharides analysed in the course of this work

Compound	Monosaccharide	Monoisotopic mass	2-AB	2-AB Me	Aniline	Aniline-Me
Hexose	Glc, Gal, Man	181.06	301.13	413.26	258.12	342.22
HexNAc	GlcNAc, GalNAc, ManNAc	222.08	342.16	454.29	299.14	383.32
Deoxyhexose	Fuc	165.06	285.14	383.25	242.12	312.12
Pentose	Xyl	151.13	271.12	369.23	228.11	298.12
N-Acetyl neuraminic acid	NeuAc	310.10	-	-	387.16	498.13

2-AB – 2 aminobenzamide derivatised monosaccharide
2-AB-Me – 2-aminobenzamide permethylated monosaccharides
Anile-Me – Aniline permethylated monosaccharide

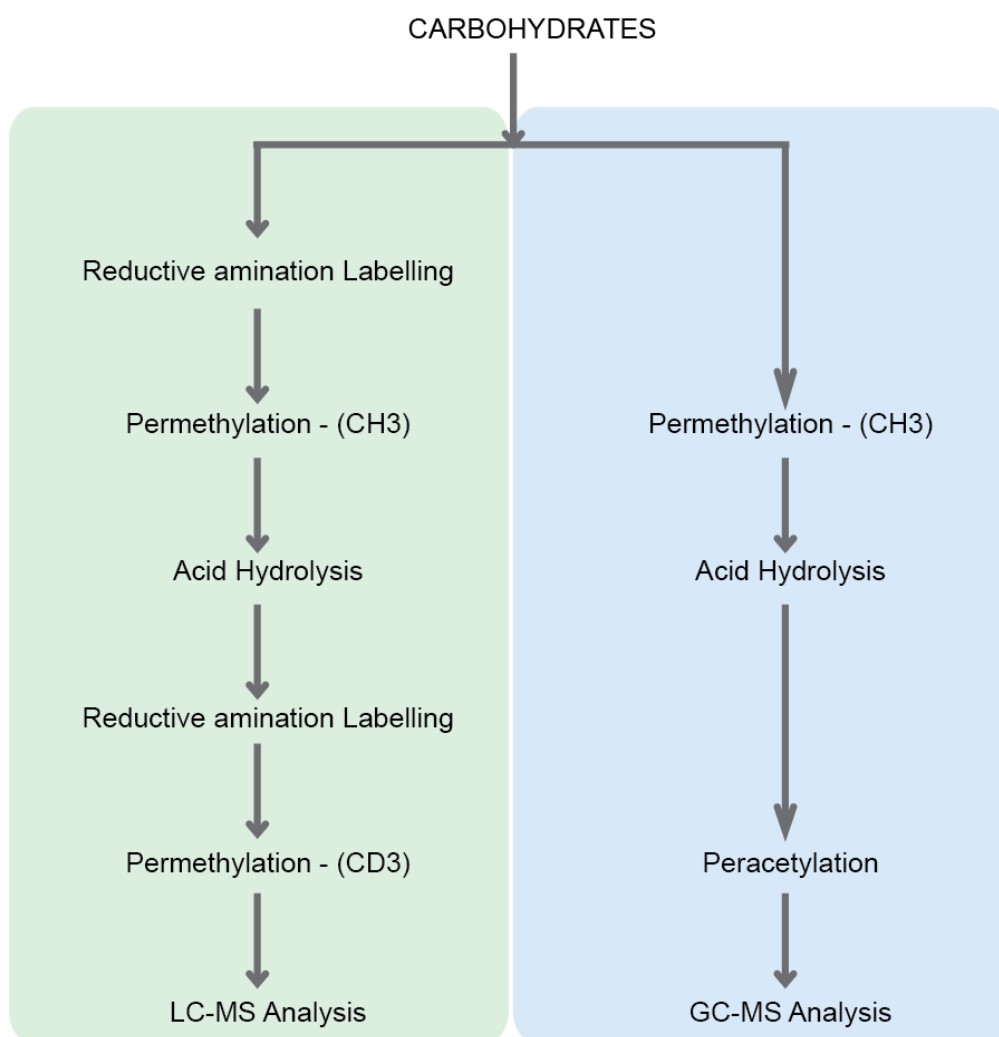


Figure 3.1: Comparison of traditional GC-MS linkage analysis (right) versus the proposed strategy for linkage analysis using LC-MS developed in the course of this work (left). A detailed optimised step-by-step procedure for monosaccharide composition and linkage analysis using LC-MS/MS is provided in the section 3.4..

3.3.2. Stereochemistry defines the fragmentation of 2-AB labelled, Permethylated monosaccharides

Offline MS analysis of the 2-AB labelled, permethylated (2-AB-Me) Hex and HexNAc molecules resulted in the preferential formation of protonated $[M+H]^+$ ions, whereas Fuc and Xyl were detected as both, protonated and sodiated ion species. For these initial MS/MS analyses the $[M+H]^+$ ions were selected and the fragmentation energies manually varied to determine the optimal fragmentation conditions that provided maximum information. The MS/MS spectra of all the three investigated Hex produced comparable fragmentation patterns, but showed distinct fragmentation signatures in the fragment ions intensities that were specific for each compound analysed (Figure 3.2). A successive loss of CH_3OH from the m/z 413.3 precursor ion produced

the observed m/z 381.2; 349.2; 317.2; 285.2 product ions. The respective monosaccharides stereochemistry obviously attributed towards this phenomenon by dictating the fragmentation route.

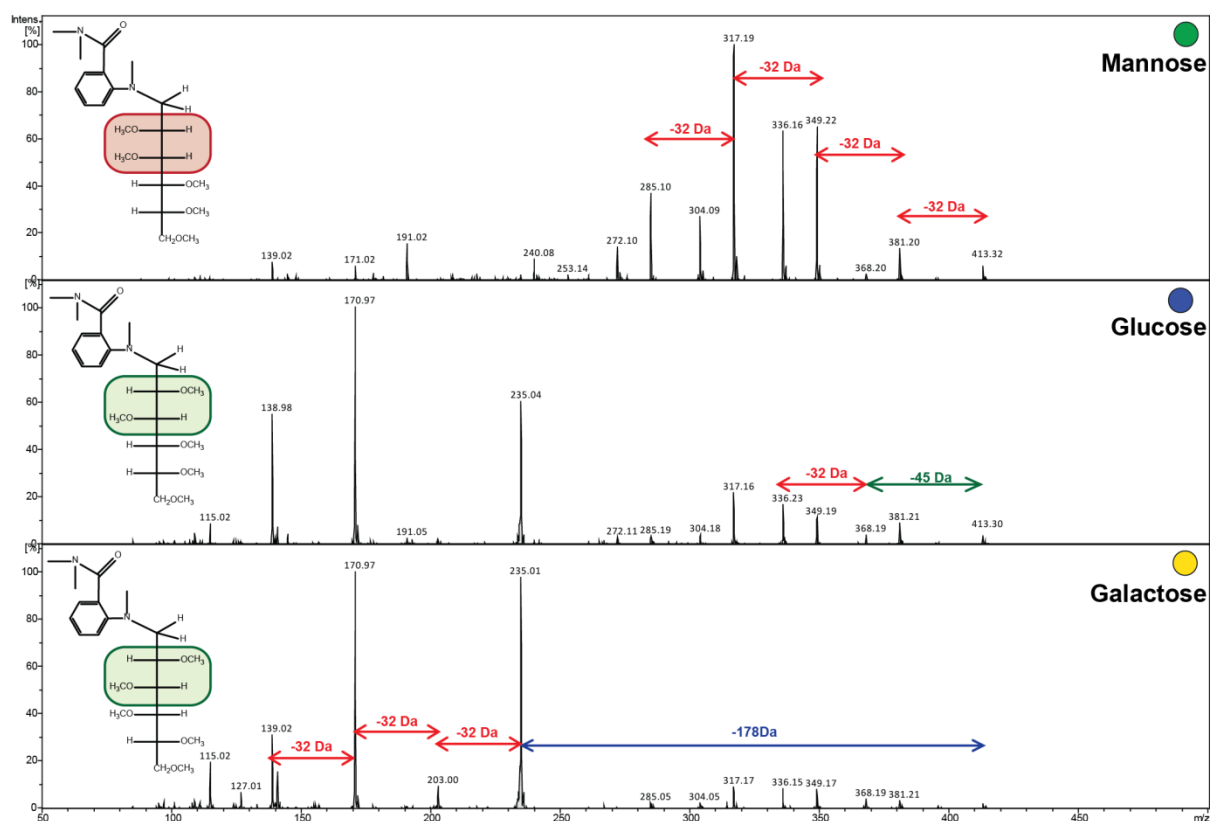


Figure 3.2: Fragmentation pattern of 2-AB-Me hexoses of the protonated precursor ion m/z 413.3. Product ions observed at m/z 381.2; 349.2; 317.2; 285.2 are the result of a successive loss of CH_3OH from m/z 413.3. The cleavage of the C–N bond between the sugar and fluorescent tag is defined by the sugar stereochemistry.

In comparison to the hexoses the MS/MS spectra of 2-AB-Me GlcNAc were rather simple. The product ion at m/z 276.2 deriving from the m/z 454.2 precursor followed a similar fragmentation as shown in Figure 3.2. Differently to the 2-AB-Me hexoses the MS/MS spectra of all three analysed 2-AB-Me HexNAc's were even more similar and comparably information poor. Despite this these fragmentation spectra still provided sufficient means to better differentiate these compounds from other low-molecular weight compounds.

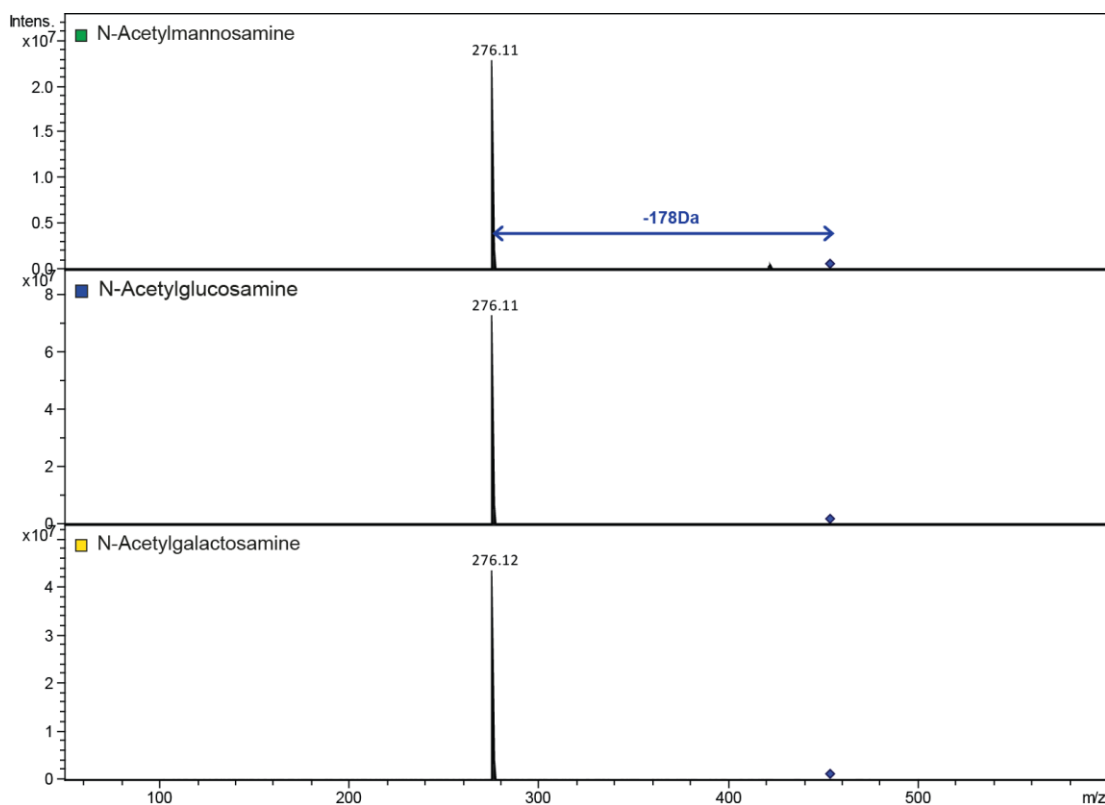


Figure 3.2: Fragmentation pattern of 2-AB Me N-acetyl hexosamines.

The MS/MS spectra obtained for 2-AB-Me fucose and xylose exhibited fragmentation signatures similar to the hexoses exhibiting a successive loss of 32 Da from the protonated precursor ion (Figure 3.3). The loss of 178.1 Da from the protonated precursor ion was consistent for all investigated 2-AB-Me derivatised monosaccharides (Scheme 3.1). The sodiated molecular ions followed the same routes albeit larger fragmentation energies were required to achieve sufficient fragmentation (data not shown).

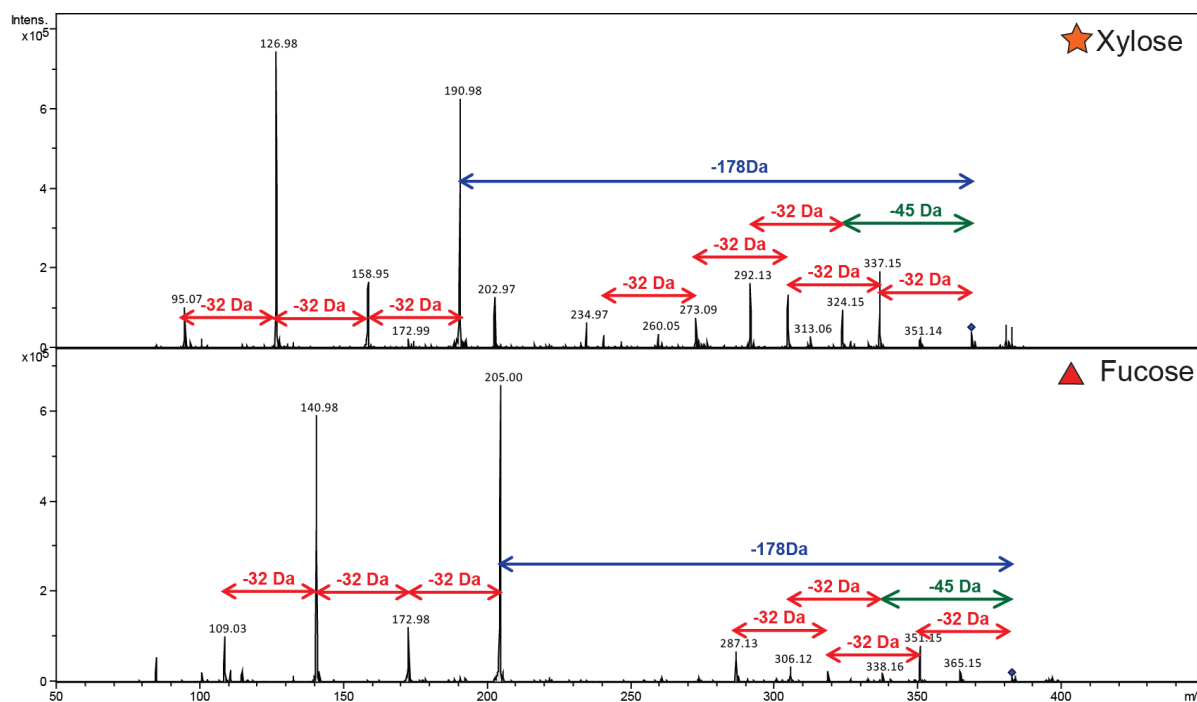
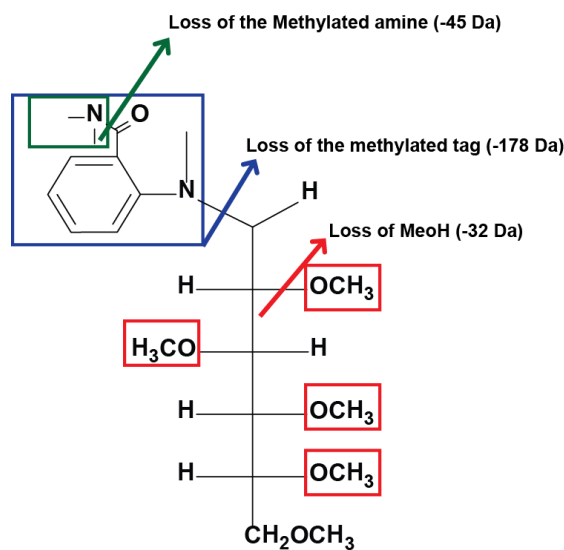


Figure 3.3: Fragmentation pattern of 2-AB-Me xylose and fucose.



Scheme 3.1: Fragmentation of 2-AB Me derivatised monosaccharides during electrospray ionisation. Primary fragments occur between the C1 atom and the Nitrogen atom of the fluorescent tag, which results in the mass loss of 178 Da from the precursor. Secondary fragmentation results in the loss of methanol (32 Da) and methylated amine (45 Da) from the precursor mass.

3.3.3. Chromatographic separation of 2-AB-ME labelled monosaccharides

As mentioned, most monosaccharides present in mammalian glycoconjugates are stereoisomers exhibiting the very same mass (Table 3.2) which cannot be discriminated by conventional MS analysers unless these isomers are separated prior MS detection. The ability to separate different isobaric, 2-AB-ME labelled monosaccharides by liquid chromatography was tested on different stationary phases. PGC is well known for its particular ability to separate unmodified oligosaccharides (22), however it was unable to provide required baseline separation of various 2-AB-Me hexoses tested (data not shown). The use of C18 reversed phase stationary phases, however, resulted in the intended separation. Optimising gradient and solvent conditions allowed baseline separation of the various tested isobaric monosaccharides (Figure 3.4).

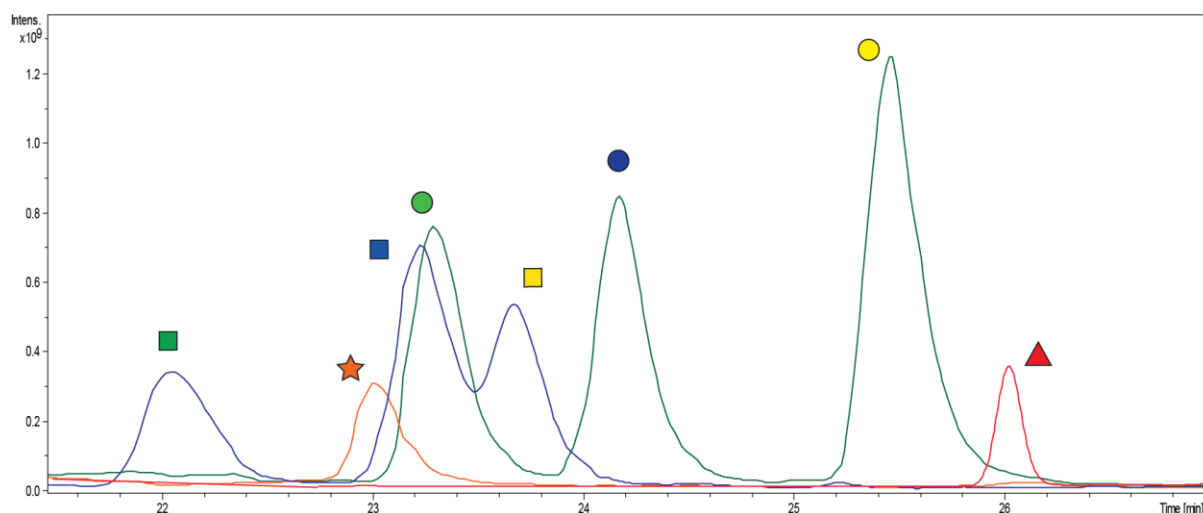


Figure 3.4: LC-ESI-MS/MS analysis of various 2-AB-Me derivatised monosaccharides separated using RP-C18 chromatography. Calculated 50 pmol of the respective compounds were injected onto the column. Extracted Ion Chromatogram (EIC) traces are shown of various monosaccharides separated under conditions described in material and methods section in the section 3.2.3.

3.3.4. Evaluation of the linkage analysis strategy using reference oligosaccharides

The successful establishment of an analysis workflow suitable for 2-AB-Me labelled monosaccharides was the basis for further developing a linkage analysis strategy (Figure 3.5). This strategy was based on the principle that first the reducing ends of the target oligosaccharides were derivatised by reductive amination with 2-AB. In the next step the free hydroxyl groups were permethylated by CH_3I , which also resulted in the methylation of any amino groups. These 2-AB-Me modified oligosaccharides were then in the next step subjected to TFA mediated acid hydrolysis, followed by a second reductive amination with 2-AB of the free reducing ends of the partially methylated monosaccharides that were produced as a consequence of the acid hydrolysis. The second permethylation step was now using CD_3I , which allowed the introduction of a CD_3 label on all C atoms that were previously involved in glycosidic linkages and thus not permethylated during the first permethylation reaction. Permethylation using CD_3 resulted in an additional mass increase of +3 Da for the presence of a single glycosidic linkage.

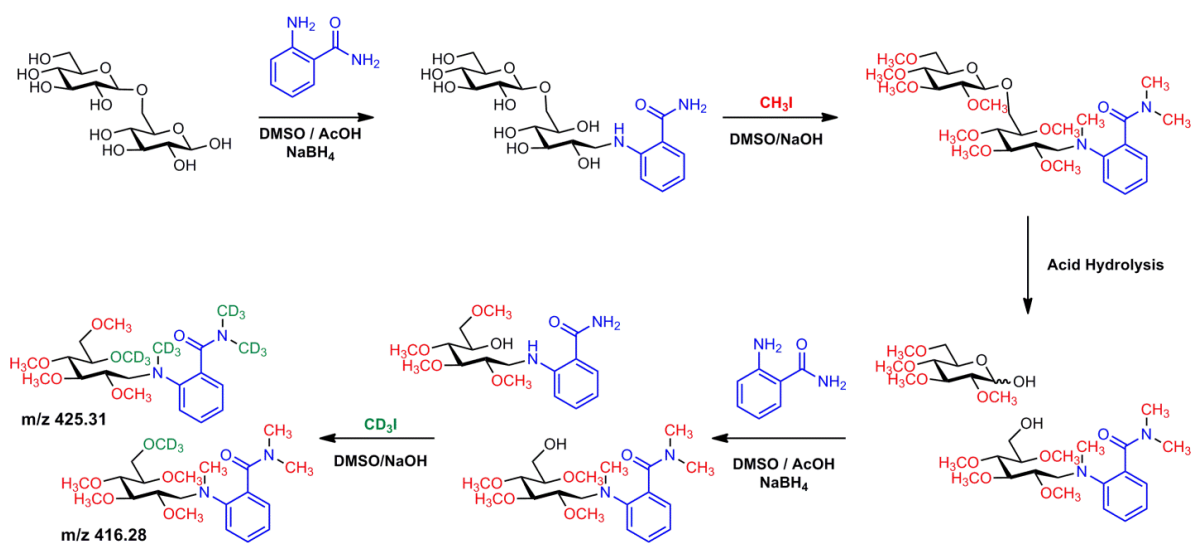


Figure 3.5: Linkage analysis workflow principle explained using gentiobiose: 1-6 linked glucose disaccharide: Carefully designed sequential permethylation and reductive amination steps prior and after acid hydrolysis enable monosaccharide identification and linkage characterisation based on specific signature MS/MS spectra (refer Figure 3.6).

A proof of principle evaluation of this strategy was performed on various reference oligosaccharide standards. Gentiobiose, lactose, trehalose, and maltoheptose were first tested before the approach was also applied on complex *N*-glycan standards. For the gentiobiose example this meant that the reducing end and non-reducing end glucose could therefore be differentiated just by their molecular masses, as these differed as a consequence of this specific labelling/permethylation strategy (Figure 3.6-A). In addition, all linkage information could be deduced from the individual fragmentation patterns (Figure 3.6-B).

2-AB-Me monosaccharides exhibited successive 32 Da losses from the precursor mass (Fragmentation route 1, scheme 3.1). In the presence of CD₃, however, 35 Da losses were detected at the linkage position, providing sufficient information for assigning the respective C-atom involved in oligosaccharide formation. Signals derived from the proposed fragmentation route 1 provided information on the reducing and non-reducing end monosaccharides, respectively (Figure 3.6-B). In the presence of CD₃ a slight shift towards earlier retention times was found compared to the respective CH₃ methylated monosaccharides (Figure 3.6-A). Such retention time shifts have also been described earlier for deuterated peptides and were associated to differences in the size of hydrogen and deuterium atoms and their binding interaction energies to the stationary phase (23). This phenomenon, however, did not alter the base line separation of the various monosaccharide stereoisomers and thus did not affect the here developed linkage and monosaccharide analysis approach. Linkage analysis results obtained from lactose is shown in Figure 3.7 differentiation glucose moiety involved in various linkages present in the standard disaccharides trehalose, lactose and gentiobiose is shown in Figure 3.8.

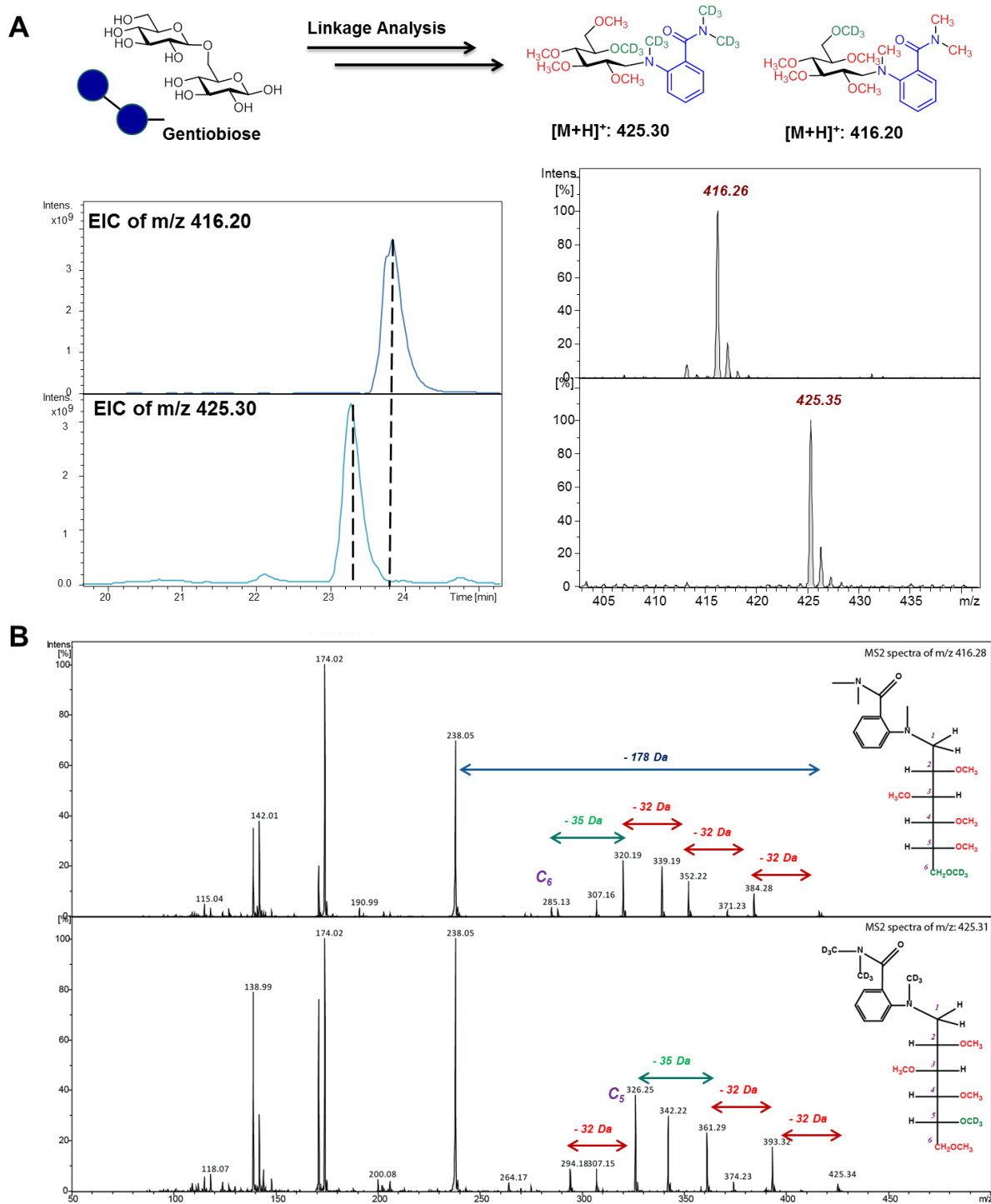


Figure 3.6: LC-ESI-MS/MS Analysis of 2-AB methylated monosaccharides derived from the standard disaccharide gentiobiose after linkage analysis (A) simplified scheme of linkage analysis and EIC of the 2-AB-Me derivatised monosaccharides obtained after linkage analysis (B) MS/MS spectra of the derived 2-AB-Me monosaccharides.

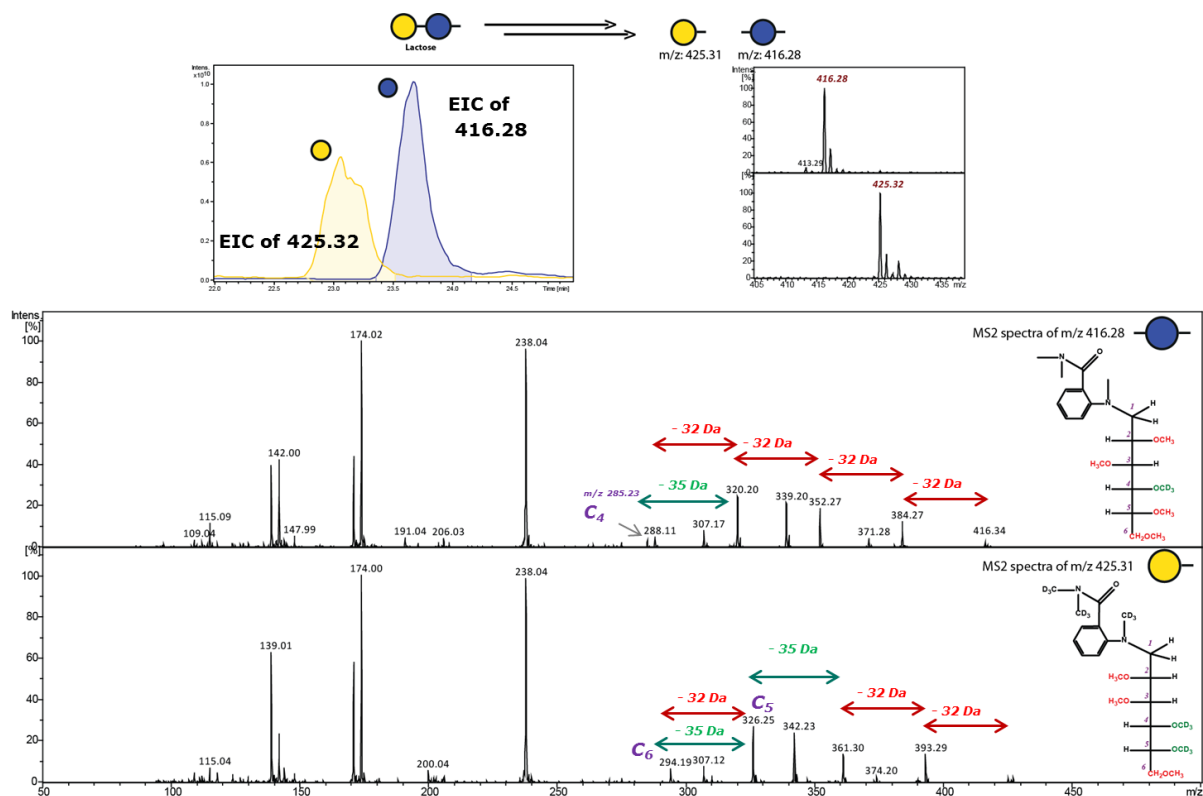


Figure 3.7: LC-ESI-MS/MS analysis of 2-AB methylated glucose and galactose derived from lactose after linkage analysis.

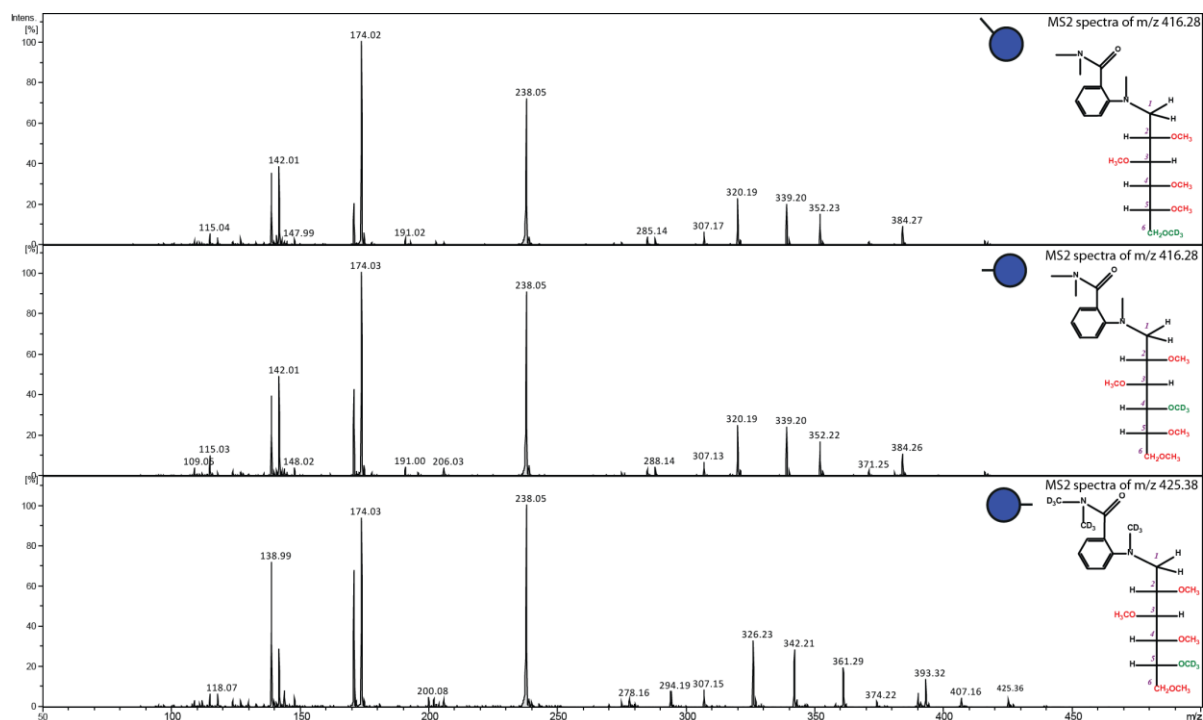
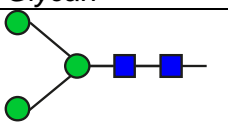
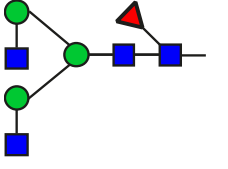
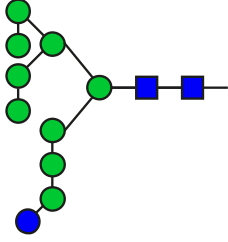


Figure 3.8: Comparison of 1-1 vs 1-4 vs 1-6 linked glucose residues derived from trehalose, lactose and gentiobiose. The linkage positions of the various glucose moieties were inferred from the sequential loss of either 32 Da or 35 Da from the precursor mass in the MS/MS spectra.

3.3.5. Method Application to N-glycans

The applicability of this approach to also be performed on protein released glycans was also tested using a panel of commercially available synthetic *N*-glycan standards (Table 3.3). Approximately 200 fmol of the reduced *N*-glycan alditols were subjected to linkage analysis. In the case of the high-mannose type *N*-glycan carrying an additional glucose residue on the 3-arm two different non-reducing end hexoses are obtained, one glucose and two mannose residues per molecule (m/z 425.35). In addition, five mannose residues that are extended on the C2 position (m/z 428.34) and two mannose residues that are extended on the C3 and C6 position (m/z 431.34) will derive from this analysis per *N*-glycan molecule (Figure 3.9). The linkage analysis results for the other standard *N*-glycan is shown in the supplementary Figure 3.10. This shows that this in the course of this thesis developed technique can be easily applied to obtained basic linkage and monosaccharide identity information even from sub-picomol amounts of starting material.

Table 3.3: List of *N*-glycans standards used to evaluate the developed linkage analysis method.

<i>Glycan Id</i>	<i>Glycan</i>	<i>Name</i>
1		M3
2		GnGnF
3		M9 Glu1

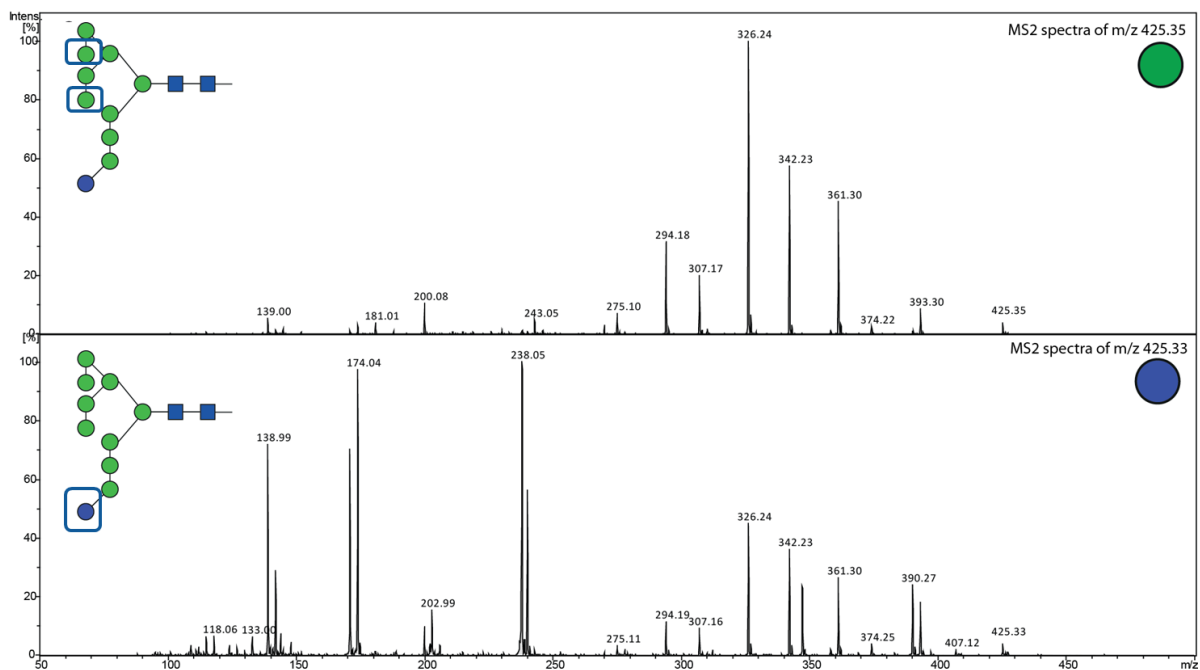


Figure 3.9: Fragment spectra of terminal mannose and glucose monosaccharides present in the Man9-Glc1 N-glycan. Linkage positions of terminal monosaccharides were identified based on the molecular mass, order of elution and the signature fragmentation pattern of 2-AB-Me monosaccharides.

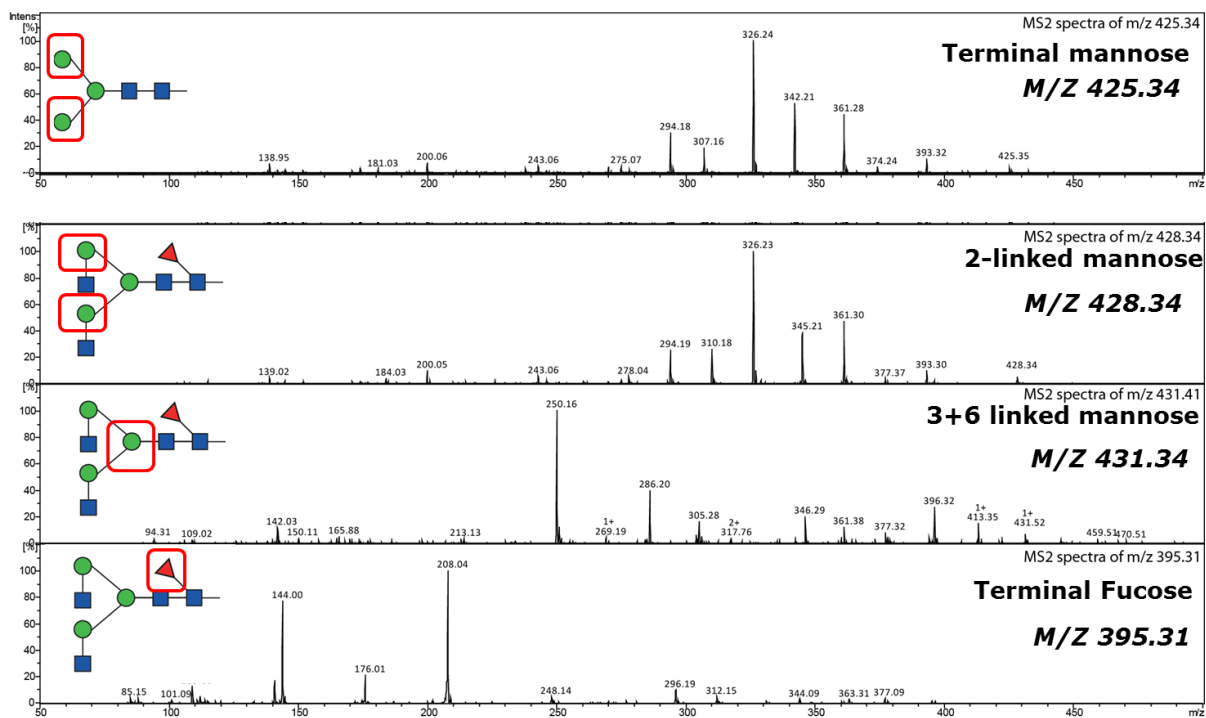


Figure 3.10: Fragment spectra of various monosaccharide residues present in the standard N-glycans listed in the table 2.

Any quantitative ratios of the monosaccharides present could be calculated by spiking in known quantities of pure monosaccharides as an internal reference compounds. The feasibility of this approach was evaluated using a maltoheptose standard that was spiked with a known quantity of a monosaccharide mix and subjected to linkage analysis (Figure 3.11).

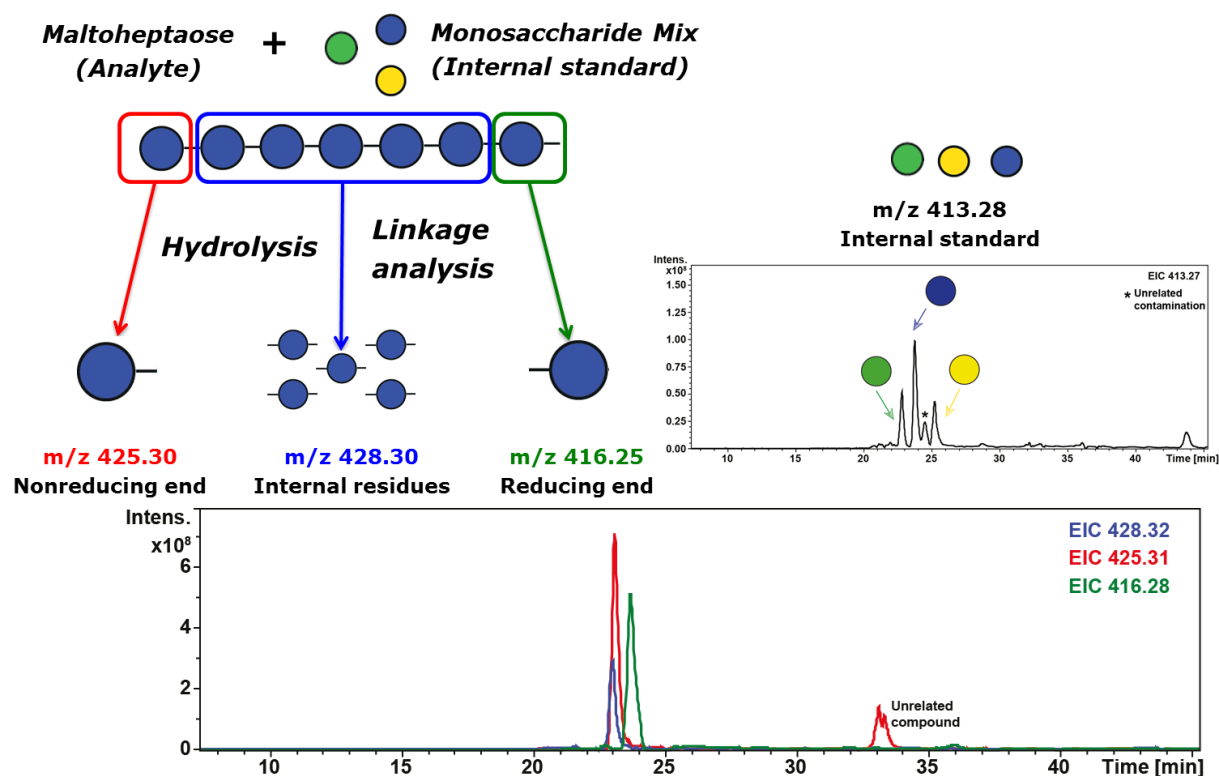


Figure 3.11: Quantification using internal standards. Known quantities of the standard monosaccharide were mixed with the analyte and subjected to monosaccharide linkage analysis. The presence of CD_3 methylation present in the analyte derived monosaccharides could be used to discriminate analyte from standards monosaccharides, as the latter only contained CH_3 methylation. The added monosaccharide served both as internal quantitation standard and for compound identification during the analysis. MS/MS spectra of the 2-AB-Me labelled monosaccharide derived from the analyte maltoheptose are shown in the figure 3.12.

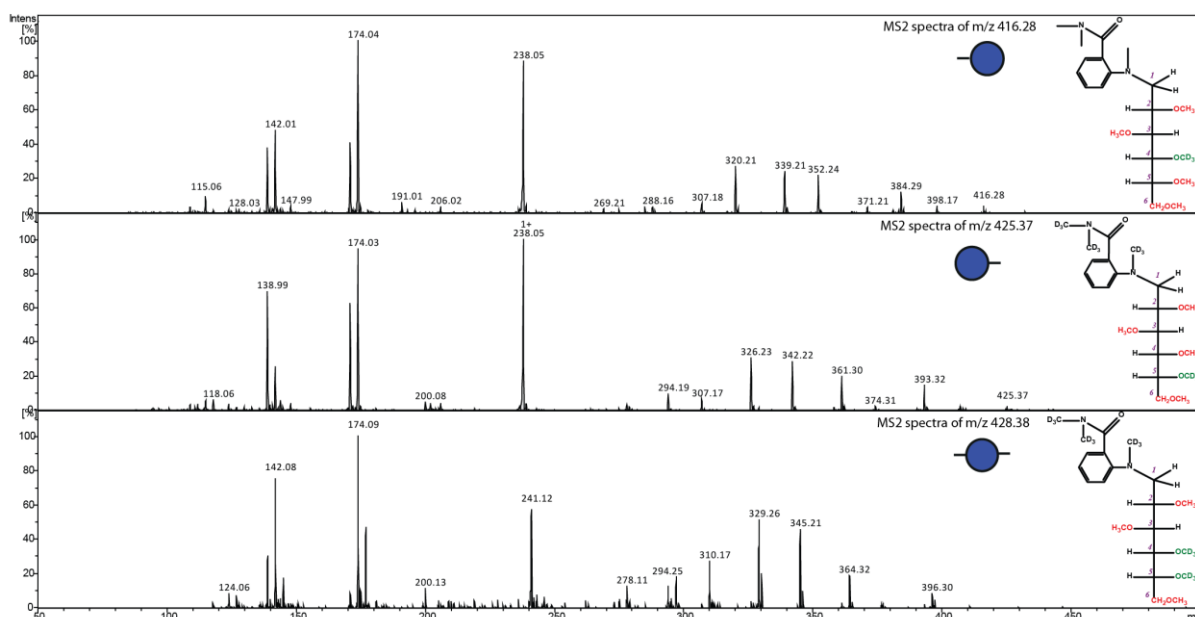


Figure 3.12: Fragment spectra of reducing end, non-reducing end and internal glucose residues present in the maltoheptose. Linkage positions of the respective monosaccharides were identified based upon the molecular mass, order of elution and fragmentation spectra.

3.3.6. Including sialic acids in the monosaccharide analysis equation

The developed approach provided the ability to identify the different monosaccharides and their linkage(s), except for sialic acids as these cannot be labelled by standard reductive amination using 2-AB. DMB (1,2-diamino-4,5 methylenedioxybenzene.2HCl) is a common labelling reagent for sialic acids, which forms a covalent bond via amination cyclisation reaction (24), but this reagent cannot be used to label other aldo-monosaccharides. Searching for a more universally applicable labelling reagent that is suited for both, keto and aldo-monosaccharides, Aniline was evaluated due to its high nucleophilic properties in comparison to 2-AB. With Aniline aldo- and keto monosaccharides could be labelled on the reducing end. The fragmentation behaviour of all analysed Aniline-Me labelled monosaccharides was also similar to that of the 2-AB-Me ones (Figure 3.13). The consolidated fragmentation pattern of Aniline-Me monosaccharide is shown in Scheme 3.2.

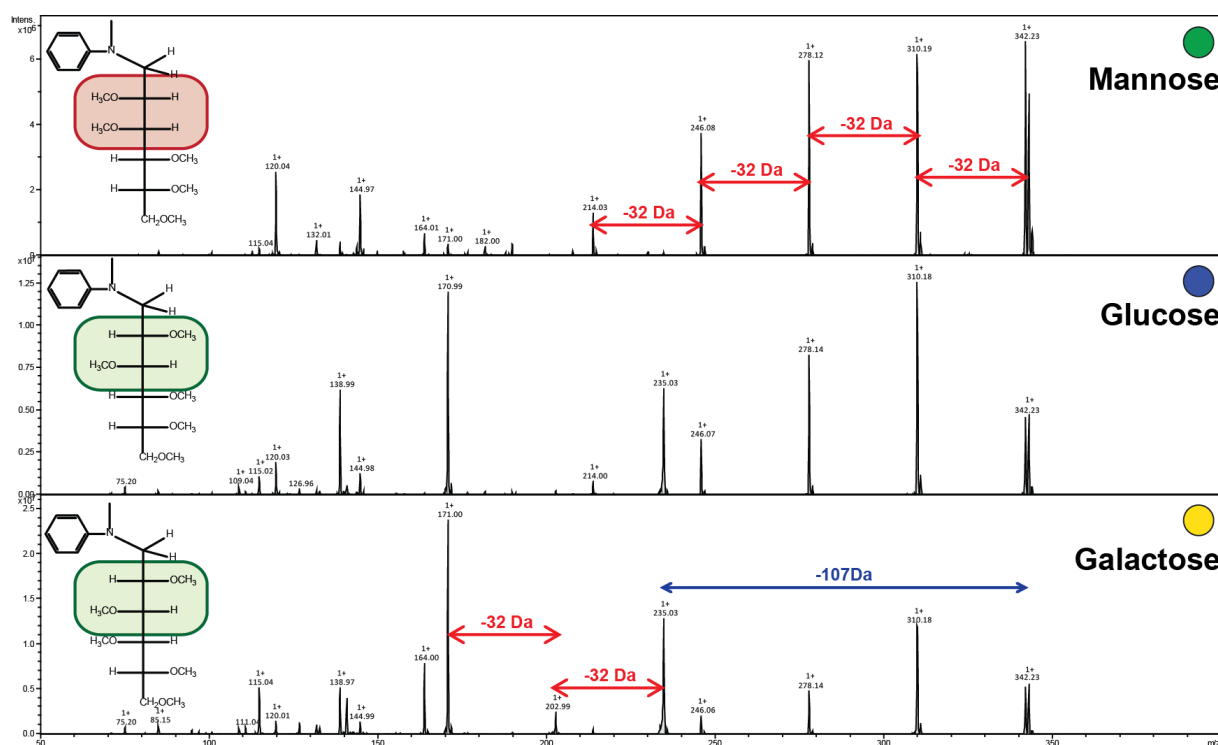
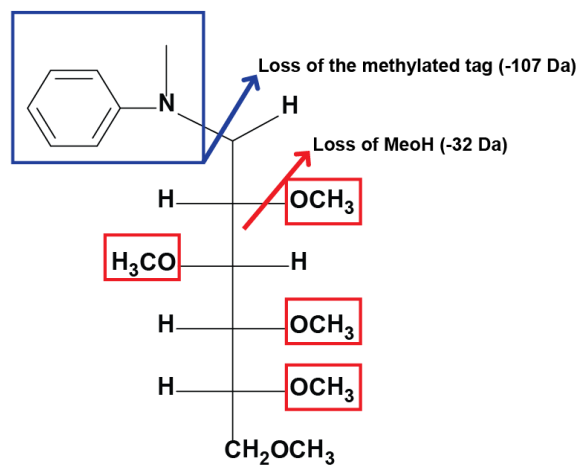


Figure 3.13: Fragmentation pattern of various Aniline-Me hexoses of the protonated precursor ion m/z 342.23. As described for 2-Ab-Me hexose, the fragmentation behaviour depends upon the stereochemistry of the monosaccharide.



Scheme 3.2: Fragmentation of Anile-Me derivatised monosaccharides during electrospray ionisation. Primary fragments occur between the C1 atom and the Nitrogen atom of the fluorescent tag, which results in the mass loss of 107 Da from the precursor. Secondary fragmentation results in the loss of methanol (32 Da) from the precursor mass.

The separation properties of the Aniline-Me derivatised monosaccharides on RP-C18 chromatography were also comparable and baseline separation could easily be achieved for isobaric monosaccharides (Figure 3.14). Interestingly, we observed that the Aniline-Me derivatised *N*-Acetylneuraminic acid eluted in two peaks. Depending on the protonation during the reductive amination, aniline can react with the C1 atom of the *N*-Acetylneuraminic acids to form both, R and S isomers. These isomers could be separated by this approach (Figure 3.15).

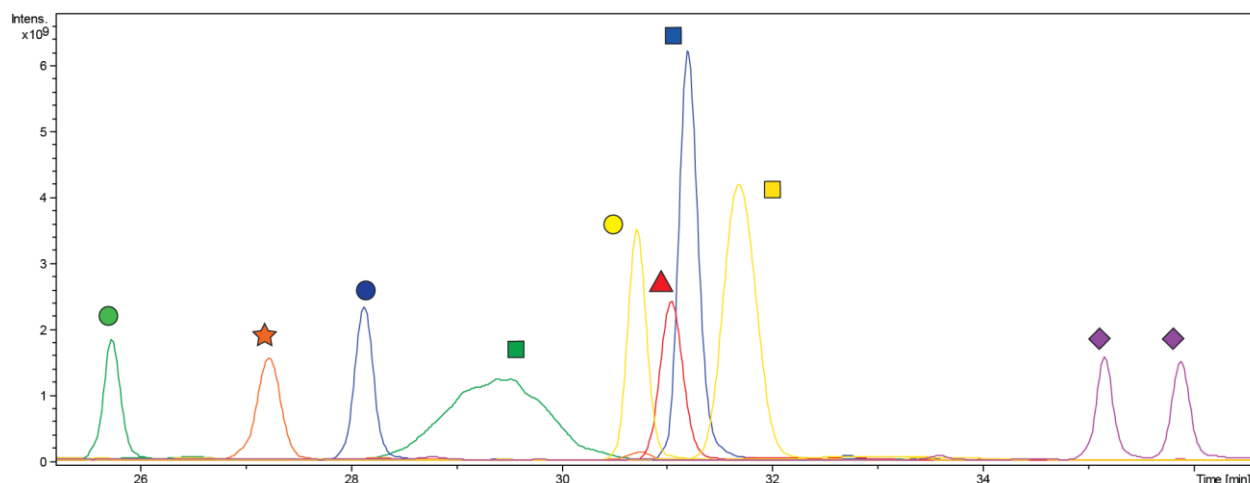


Figure 3.14: LC-ESI-MSMS analysis of various Aniline-Me derivatised monosaccharides separated using RP-C18 chromatography. Calculated 50 pmol of the respective compounds were injected onto the column. Extracted Ion Chromatogram (EIC) traces of various monosaccharides separated under optimised conditions are overlaid.

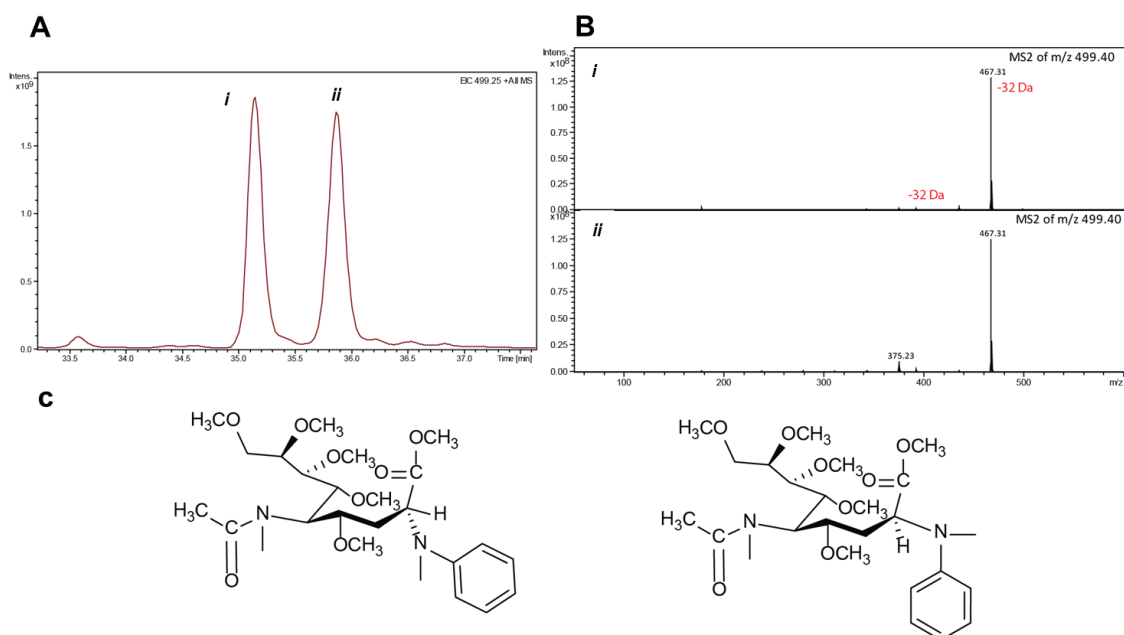


Figure 3.15: N-acetylneuraminic acid labelling with aniline. (A) EIC-trace of Aniline-Me derivatised N-acetylneuraminic acid eluting at two different time points. (B) MSMS spectra of m/z 499.40 eluting at the two positions having identical fragmentation spectra. Here we observe the characteristic 32 Da loss from the precursor mass for the aniline-Me derivatised monosaccharides. (C) Chemical structure of R and S isomers of Aniline-Me N-acetylneuraminic acid resolved by RP chromatography.

In addition, the suitability of aniline as derivatisation agent for the linkage analysis was tested using the standard disaccharides gentiobiose and Trehalose and the results are shown in Figure 16. These initial results indicated that both aniline as well as 2-AB can be employed for linkage analysis providing base line separation of various stereoisomers. Nevertheless, aniline provides an additional advantage of being a universal derivatisation agent able to label both aldo and keto-sugar.

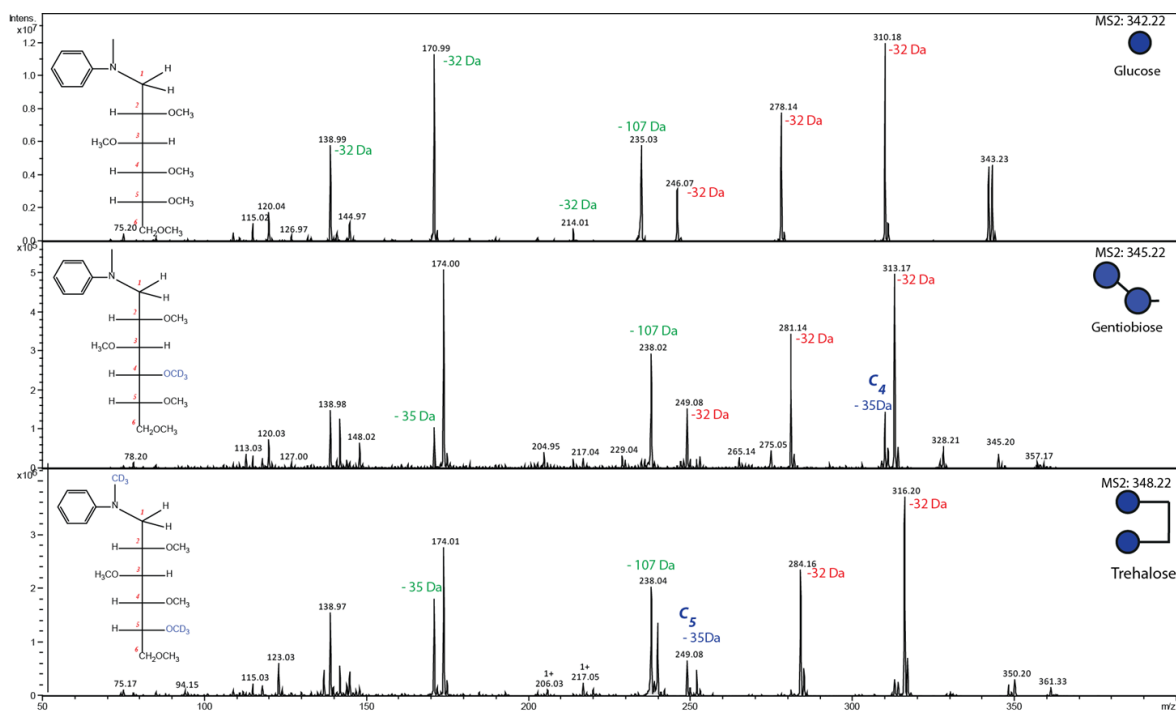


Figure 3.16: Linkage analysis using Aniline. Comparison of fragmentation pattern of glucose vs 1-1 vs 1-4 linked glucose residues derived from gentiobiose and trehalose.

However, one point of concern in the developed method for linkage analysis is that *N*-acetylhexosamine was not detected due to some unexplainable reason. However, the developed approach is still a valuable tool in providing glycan branching point information for each arm from a minimal sample amount.

3.4 OPTIMISED STEP-WISE PROCEDURE FOR MONOSACCHARIDE LINKAGE AND COMPOSITIONAL ANALYSIS

A) Monosaccharide derivatisation

Reducing end derivatisation

Labelling Reagent 1: 0.35 M 2-Aminobenzamide and 1 M sodium cyanoborohydride were freshly prepared in dimethyl sulfoxide:acetic acid (7:3, v/v).

Labelling Reagent 2: 0.35 M Aniline and 1 M sodium cyanoborohydride were freshly prepared in dimethyl sulfoxide:acetic acid (7:3, v/v).

1. Aliquot and freeze-dry the glycan sample in an Eppendorf tube
2. Add 50 μ L of the labelling reagent 1 or 2.
3. Incubate for 120 min at 65°C.

- Cool down to room temperature.
- Add 500 μL of acetone or DCM and vortex for 30 s.
- Centrifuge at 13,000 rpm for 5 min, the entire solvent was decanted and the procedure was repeated twice
- Purify the derivatised carbohydrates using a carbon solid-phase extraction tip.

B) CH_3I permethylation

- Dry the glycan sample in glass tube and cap with a Teflon-lined screw cap.
- Critical step: Samples must be dry and free of salts; otherwise, they will not completely methylate.
- Add 500 μL of DMSO, cap and sonicate the tubes for 20–30 min.
- Add approximately 50 mg of the powdered NaOH to each sample
- Act quickly and do not attempt to weigh NaOH pellets, as they will absorb water rapidly.
- Take care to use an even suspension of NaOH and to drop the reagent straight onto the sample—do not get the NaOH suspension on the side of the tube, or the reagent will be limiting.
- Cap and sonicate for 20–50 min.
- Add 30 μL of CH_3I and sonicate for 10 min.
- Caution: Only use CH_3I in a fume hood, as it is extremely toxic and a suspected carcinogen.
- Add 20 μL of CH_3I and sonicate for 10 min.
- Add 50 μL of CH_3I and sonicate for 20 min.
- Add 500 μL of water and 500 μL of DCM. Cap and vortex well (>40 s per sample). Centrifuge briefly to separate the phases.
- Remove and discard the aqueous (upper) phase. Wash the lower DCM phase three times with 500 μL water.
- Dry the lower DCM phase with a stream of dry nitrogen.

C) Acid hydrolysis

- Add 200 μL of 2.0 M TFA to the methylated sample. Hydrolyse for 90 min at 121°C in a fan-forced oven.
- Cool the sample and place the tube in a warm water bath (~30°C) and evaporate to dryness with a stream of nitrogen.

24. Alternatively samples can be transferred to an Eppendorf tube and dried in vacuo using a SpeedVac concentrator.
25. Add 100 μ L of acetonitrile to the samples and dry the samples in the SpeedVac concentrator. Repeat this procedure three or five times to remove any residual TFA.
26. Purify the derivatised carbohydrates using a C18 solid-phase extraction tip.
27. Reductive amination derivatisation – 2AB

This is done exactly as described in the steps 1-4.

28. Derivatised samples can now be extracted via DCM: Water extraction. Lower DCM phase contains the derivatised glycan.

D) CD₃I Permethylation

This is done exactly as described in the previous section 8-17 except that deuterated methyl iodide (CD₃I) is used instead.

3.5 CONCLUSION

The in the course of this work developed method provided an easily adaptable linkage analysis approach that made use of a LC-ESI-MSMS equipment available in the majority of proteomics labs. Identification and in-depth characterisation of glycans also requires tools for the sensitive and selective analysis of monosaccharides and orthogonal means of linkage characterisation. These methods are crucial to profile complex glycoconjugates, in particular if glycosylation in less well studied organisms needs to be determined. Traditional methods used for compositional and/or linkage monosaccharide analysis such as HPAE-PAD or GC-MS require dedicated instrumentation, which is not widespread available. Here we present a simple, sensitive and easily adaptable method suitable for the unambiguous identification of most commonly occurring monosaccharides including N-Acetylneuraminic acids. Using the power of C18-RP chromatographic separation combined with mass spectrometric detection reductively labelled and permethylated monosaccharides could be identified and the respective linkages they were involved in determined. The here reported method was determined to show a limit of detection of 500 fmol for all analysed monosaccharides. The method herein reported was sufficiently sensitive to provide useful results from as low as 10-20 pmol of initial analyte when just compositional monosaccharide analysis was required. For a comprehensive analysis

also including linkage determination ≥ 200 pmol initial analyte was required, as this required multiple derivatisation steps that naturally resulted in larger sample losses. The here presented approach provides one important step forward towards an unambiguous assignment of monosaccharide identity and linkage beyond inferring biosynthetic knowledge. This is of particular importance for less well studied organisms such as bacteria that use a much wider range of the monosaccharide space compared to eukaryotic organisms.

3.6 REFERENCES

1. Alley WR, Novotny MV. Structural Glycomic Analyses at High Sensitivity: A Decade of Progress. *Annu Rev Anal Chem.* 2013;6:237-65.
2. Marino K, Bones J, Kattla JJ, Rudd PM. A systematic approach to protein glycosylation analysis: a path through the maze. *Nat Chem Biol.* 2010;6(10):713-23.
3. Maley F, Trimble RB, Tarentino AL, Plummer TH. Characterization of Glycoproteins and Their Associated Oligosaccharides through the Use of Endoglycosidases. *Analytical Biochemistry.* 1989;180(2):195-204.
4. Guttman A. Multistructure sequencing of N-linked fetuin glycans by capillary gel electrophoresis and enzyme matrix digestion. *Electrophoresis.* 1997;18(7):1136-41.
5. Björndal H, Lindberg B, Svensson S. Mass spectrometry of partially methylated alditol acetates. *Carbohyd Res.* 1967;5(4):433-40.
6. Kochetkov NK, Chizhov OS, Zolotare.Bm, Wulfson NS. Mass Spectrometry of Carbohydrate Derivatives. *Tetrahedron.* 1963;19(12):2209-&.
7. Fournet B, Strecker G, Leroy Y, Montreuil J. Gas-Liquid-Chromatography and Mass-Spectrometry of Methylated and Acetylated Methyl Glycosides - Application to the Structural-Analysis of Glycoprotein Glycans. *Anal Biochem.* 1981;116(2):489-502.
8. Rohrer JS, Basumallick L, Hurum D. High-performance anion-exchange chromatography with pulsed amperometric detection for carbohydrate analysis of glycoproteins. *Biochemistry (Mosc).* 2013;78(7):697-709.
9. Rogatsky E, Jayatillake H, Goswami G, Tomuta V, Stein D. Sensitive LC MS quantitative analysis of carbohydrates by Cs⁺ attachment. *J Am Soc Mass Spectrom.* 2005;16(11):1805-11.
10. McIntosh TS, Davis HM, Matthews DE. A liquid chromatography-mass spectrometry method to measure stable isotopic tracer enrichments of glycerol and glucose in human serum. *Anal Biochem.* 2002;300(2):163-9.
11. Wan ECH, Yu JZ. Determination of sugar compounds in atmospheric aerosols by liquid chromatography combined with positive electrospray ionization mass spectrometry. *J Chromatogr A.* 2006;1107(1-2):175-81.
12. Kato Y, Numajiri Y. Chloride Attachment Negative-Ion Mass-Spectra of Sugars by Combined Liquid-Chromatography and Atmospheric-Pressure Chemical Ionization Mass-Spectrometry. *J Chromatogr-Biomed.* 1991;562(1-2):81-97.
13. Rogatsky E, Tomuta V, Stein DT. LC/MS quantitative study of glucose by iodine attachment. *Anal Chim Acta.* 2007;591(2):155-60.
14. Mechref Y. Monosaccharide Compositional Analysis of Glycoproteins and Glycolipids: Utility in the Diagnosis/Prognosis of Diseases. *Capillary Electrophoresis of Carbohydrates: From Monosaccharides to Complex Polysaccharides.* 2011:237-67.
15. Hammad LA, Derryberry DZ, Jmeian YR, Mechref Y. Quantification of monosaccharides through multiple-reaction monitoring liquid chromatography/mass spectrometry using an aminopropyl column. *Rapid Commun Mass Sp.* 2010;24(11):1565-74.
16. Hammad LA, Saleh MM, Novotny MV, Mechref Y. Multiple-Reaction Monitoring Liquid Chromatography Mass Spectrometry for Monosaccharide Compositional Analysis of Glycoproteins. *J Am Soc Mass Spectrom.* 2009;20(6):1224-34.
17. Windwarder M, Figl R, Svehla E, Mocsai RT, Farcet JB, Staudacher E, et al. "Hypermethylation" of anthranilic acid-labeled sugars confers the selectivity required for liquid chromatography-mass spectrometry. *Anal Biochem.* 2016;514:24-31.

18. Ciucanu I, Kerek F. A Simple and Rapid Method for the Permethylation of Carbohydrates. *Carbohydr Res.* 1984;131(2):209-17.
19. Ciucanu I, Costello CE. Elimination of oxidative degradation during the per-O-methylation of carbohydrates. *J Am Chem Soc.* 2003;125(52):16213-9.
20. Bigge JC, Patel TP, Bruce JA, Goulding PN, Charles SM, Parekh RB. Nonselective and Efficient Fluorescent Labeling of Glycans Using 2-Amino Benzamide and Anthranilic Acid. *Anal Biochem.* 1995;230(2):229-38.
21. Pabst M, Kolarich D, Poltl G, Dalik T, Lubec G, Hofinger A, et al. Comparison of fluorescent labels for oligosaccharides and introduction of a new postlabeling purification method. *Anal Biochem.* 2009;384(2):263-73.
22. Stavenhagen K, Kolarich D, Wuhrer M. Clinical Glycomics Employing Graphitized Carbon Liquid Chromatography-Mass Spectrometry. *Chromatographia.* 2015;78(5-6):307-20.
23. Iyer SS, Zhang ZP, Kellogg GE, Karnes HT. Evaluation of deuterium isotope effects in normal-phase LC-MS-MS separations using a molecular modeling approach. *J Chromatogr Sci.* 2004;42(7):383-7.
24. Hara S, Takemori Y, Yamaguchi M, Nakamura M, Ohkura Y. Fluorometric high-performance liquid chromatography of N-acetyl- and N-glycolylneuraminic acids and its application to their microdetermination in human and animal sera, glycoproteins, and glycolipids. *Anal Biochem.* 1987;164(1):138-45.

4. DEVELOPMENT OF FAST AND EFFICIENT STRATEGIES FOR THE PRODUCTION OF GLYCOSYLATED AMINO ACID BUILDING BLOCK

4.1. BACKGROUND

MS is frequently used in clinical proteomics to investigate protein complexes and to detect qualitative and quantitative differences in protein disease markers from readily available body fluids like serum, saliva or urine. The majority of human proteins are co- and post-translationally modified and PTM dependent modulation of protein function has attracted substantial attention in recent year especially in the case of glycosylation. Glycoproteins are ubiquitous and essential components of the plasma membrane, the extracellular matrix and are major key players in serum and secreted fluids (1). Most glycoproteins are present in multiple glycoforms (2). This heterogeneity has been described to affect protein activity, efficacy, stability, degradation, immunogenicity and solubility. Glycoproteins containing more than a single glycosylation site are frequently exhibiting different glyco-profiles on the respective glycosylation sites (3-7), and disease induced glycan-alterations can also just occur on selective positions.

The ability to accurately determine glycoprotein primary structure is a prerequisite to exploit site specific glycosylation information as disease markers or to understand the functional role of glycans present on specific sites in individual glycoproteins. This capacity, however, largely depends upon sensitive, reliable & high throughput analytical methods and data analysis workflows that allow detection, quantitation and identification of these compounds. A key feature for accurate MS analyses is the availability of particular standards that enable the user to tune and control their instrument to avoid any artefacts. Compared to proteomics counterparts, analytical methods and data analysis workflows available for glycoprotein analysis are still in its infancies, also due to the fact that synthetic glycopeptide standards are not readily accessible. Thus having a methodology in hand that enables the tailored synthesis of individual glycopeptides will first promote the development of analytical workflows and reliable software tools for glycopeptide analysis that subsequently will significantly contribute towards

understanding the biological role of glycoconjugates. Total glycopeptide synthesis requires the combinatorial approach of carbohydrate and peptide chemistry. Despite tremendous advances in carbohydrate chemistry the production of large oligosaccharides with diverse building blocks still requires substantial time and material resources, in particular if an *in vivo* like linkage between an oligosaccharide and an amino acid are required. The fact that the glycosylation machinery is considerably conserved in eukaryotic organisms, however, provides the opportunity to isolate biologically important glycan structures from various easily accessible and cheap sources. In this work a simple and environmentally friendly strategy for the isolation of milligram quantities of glycosylated amino acid precursors from egg yolk was developed that provides the crucial glycosylated building blocks necessary for step-wise solid phase glycopeptide synthesis. This building block was used to synthesise a panel of *N*-glycopeptides that subsequently were employed to optimise and evaluate (i) glycopeptide enrichment efficiency using ZIC-HILIC, (ii) the influence of the charge state on glycopeptide fragmentation, (iii) the ability of the captive spray nanoBooster to enable enrichment un-biased glycoproteomics, and (iv) provide novel possibilities aiming towards automated glycopeptide identification.

4.2. MATERIALS AND METHODS

4.2.1. MATERIALS

Standard amino acids for SPPS were purchased from Iris Biotech GmbH (Marktredwitz, Germany), and Wang ChemMatrix resin was from Novabiochem (Darmstadt, Germany). Fresh chicken eggs were obtained from the local market in Dahlem (Berlin, Germany) and the Egg Yolk Powder was obtained from Myprotein online shop (Greater Manchester, England - <http://www.myprotein.com/home.dept>).

4.2.2. HEXAPEPTIDE ISOLATION

An egg yolk hexapeptide carrying a disialylated biantennary *N*-glycan was initially prepared from egg yolks as described previously (8, 9). This method was significantly modified for isolating the glycosylated hexapeptides by precipitating the protein using chloroform/methanol method (10) as described in detail in the results section. The obtained crude product was volume reduced using rotovap and the concentrated solution subjected to either gel filtration on Sephadex G50

(fine) or Sephadex G25 (fine) (25 X 935 mm) and eluted with 100 mM Ammonium acetate (pH 7) at the flow rate of 1 mL/min at 10°C. Sixty fractions of 10 mL were collected and the resorcinol positive fractions were pooled and the obtained products were verified by off-line ESI-MS.

4.2.3. PRONASE DIGESTION

Glycopeptides were dissolved in 25 mM Tris/HCl (pH: 7.8) containing 3 mM CaCl₂ and 0.2% NaN₃. After the addition of pronase (Potein: Pronase = 13:1) (Protease from *streptomyces griseus*, P6911-1g, Lot no: SLB 2433V), the samples were incubated overnight at 37°C. The digestion products obtained were subjected to gel filtration on Sephadex G50 (fine) as described in the hexapeptide isolation section.

4.2.4. GLYCOPEPTIDE SYNTHESIS

All peptides and glycopeptides were synthesised by SPPS using previously reported fluorenylmethoxycarbonyl (Fmoc) protocols. All peptides and glycopeptides were synthesised semi-automatically using a commercially available Wang ChemMatrix[®] resin using on an automated microwave assisted peptide synthesiser Liberty Blue[©] and 5-mL or 10-mL disposable polypropylene syringes with a bottom filter. Sialic acid residues were selectively protected by esterification with benzyl bromide prior their use in SPPS (11-14). The coupling of the glycosylated Asn building blocks was performed as described by Unverzagt and co-workers (15).

4.2.5. GLYCOPEPTIDE PURIFICATION

Synthesised glycopeptides were purified using preparative C18-RP HPLC on an Agilent 1200 series ELSD HPLC system (Agilent, Santa Clara, CA). The lyophilisate was dissolved in water and subjected to C18-RP chromatography (XBridge BEH C18 column, 130 Å, 3.5 µm, 4.6 mm x 150 mm, Waters, Milford, MA) using a linear gradient (1 mL/min) from 1% to 50% acetonitrile containing 0.1% TFA within 35 min, followed by an increase to 100% over 10 min. Fractions of interest from several runs were combined and analysed by MS to confirm their composition.

4.2.5. EXOGLYCOSIDASE DIGEST

Aliquots of the purified synthetic glycopeptide carrying bi-antennary sialylated *N*-glycan were dried in vacuum and reconstituted in 50 mM sodium acetate containing 5 mM CaCl₂, pH 5.5. The reconstituted glycopeptides were then subjected to a combination of exoglycosidase digestion either in combination or sequentially [α 2-3,6,8 Neuraminidase (NEB- P0720 L), β -1-4galactosidase S (proglyAn) and β -N-Acetyl-glucosaminidase (NEB-P0744S) . All the reactions were performed at 37° overnight.

4.3 RESULTS AND DISCUSSION

4.3.1. ISOLATION OF THE GLYCOSYLATED HEXAPEPTIDE FROM FRESH EGG YOLK

4.3.1.1. Seko Approach (8)

First sialyl *N*-glycopeptides were isolated from the fresh chicken egg yolks as described by Seko *et al* previously (8). Fresh yolks from 60 eggs were deproteinised and delipidated using 10% phenol at 0°C for 3 h following centrifugation to pellet down the proteins. The aqueous phase (supernatant) containing the soluble compounds was lyophilised. The obtained concentrate was dissolved and subjected to Sephadex G-25 gel chromatography to easily separate the target glycopeptide from smaller molecular weight components. The fractions were monitored for the presence of sialic acids and the resorcinol positive fractions containing the target glycopeptide were pooled, concentrated and lyophilised (Figure 4.1). About 600 mg of the crude glycopeptide was obtained after a single chromatographic step, and the presence of the desired sialylglycopeptide was further verified by off-line ESI-MS/MS analysis (Figure 4.2).

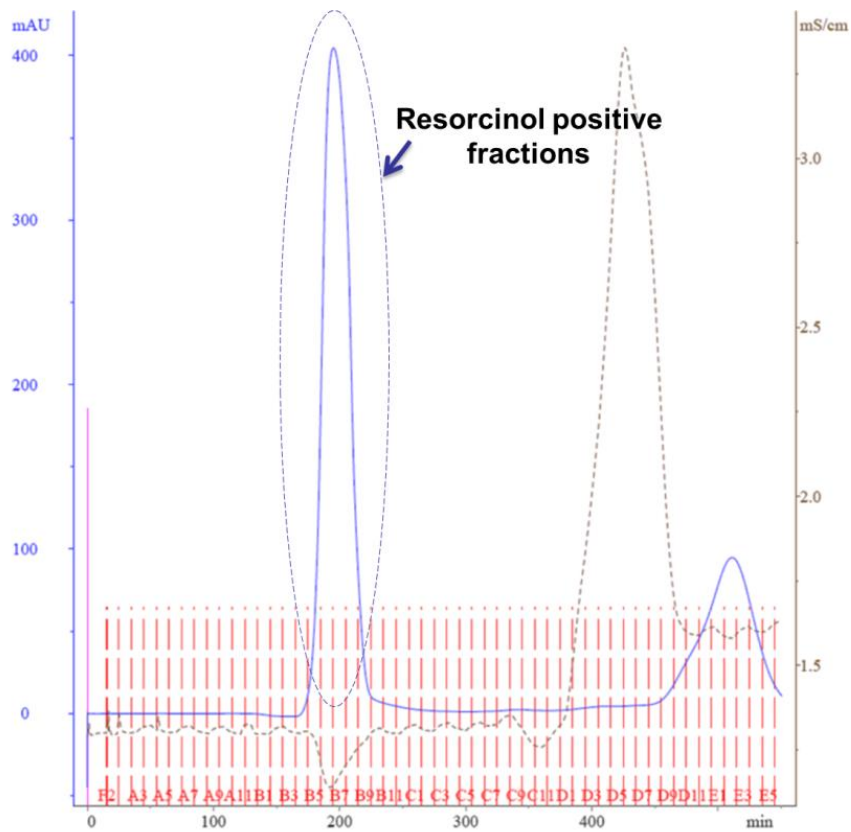


Figure 4.1: Chromatogram depicting the successful one-step isolation of the sialylglycopeptide by Sephadex G-25 gel chromatography after phenol-treatment of egg yolk. Other small molecular weight contaminants (dashed line) are clearly eluting in the later stage of the chromatography.

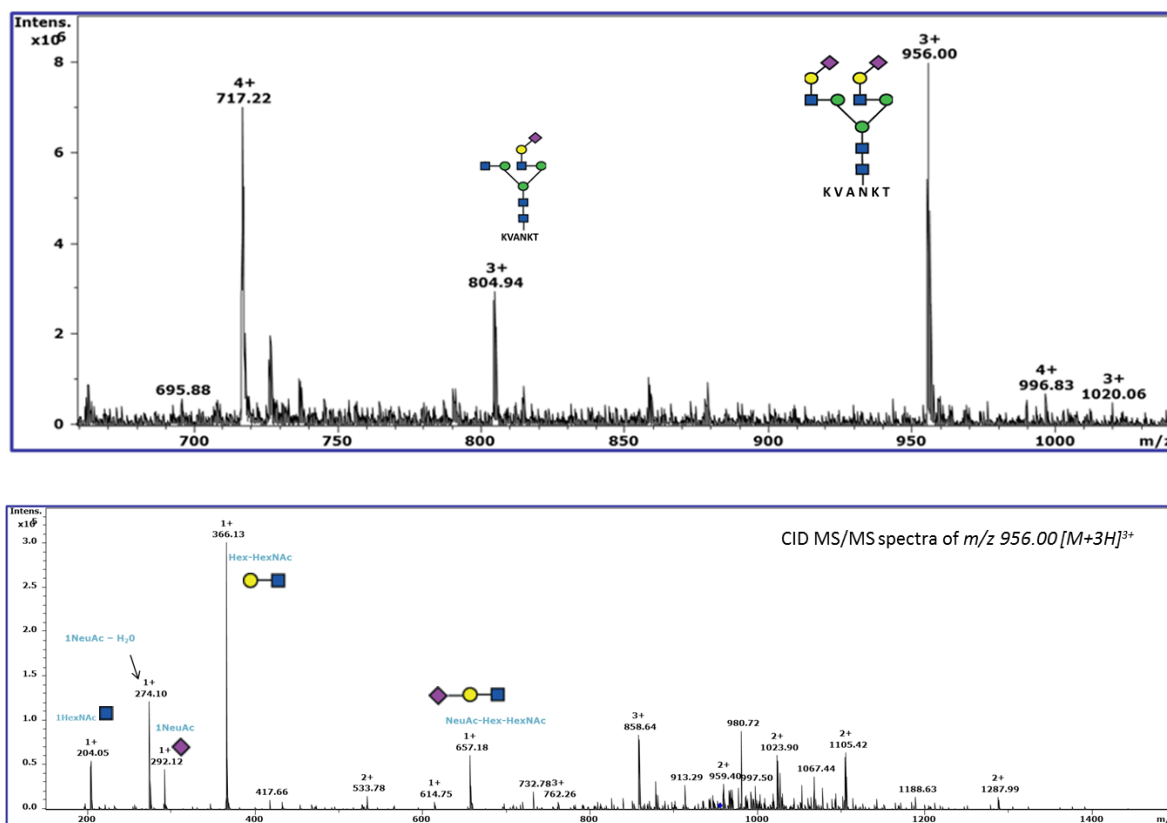


Figure 4.2: MS and MS/MS spectra of resorcinol positive fractions pooled after gel filtration analysed via offline MS. The m/z 717.22 $[M+4H]^{4+}$ and m/z 956.00 $[M+3H]^{3+}$ signals correspond to the egg yolk hexapeptide carrying a sialylated biantennary N-Glycan (top spectrum). Further, CID-MS/MS data confirmed the presence of the expected sialylated N-linked glycan (bottom spectrum).

4.3.1.1. A new approach based on the "Wessel protocol"

Despite the fact that the desired hexapeptide could be isolated in milligram scale quantities using the initially published Seko protocol, this approach came with the major drawback that large amounts of phenol were required. This represented significant downstream processing issues as phenol waste represents an environmental hazard. Furthermore, additional sources for glycosylated amino acid building blocks that possibly were present in the phenol precipitate were also lost for further enrichment. Thus a different approach was developed to overcome these issues by isolating the hexapeptide using a chloroform/methanol protocol that concomitantly precipitated proteins and removed lipids into different fractions. The desired egg yolk glycosylated hexapeptide was recovered in the aqueous layer, which was subsequently concentrated using the rotovap and lyophilised.

The glycosylated hexapeptide containing lyophilisate was subjected to Sephadex G-25 (fine) gel chromatography for further purification (Figure 4.3). Resorcinol

positive fractions were pooled, lyophilised (4.5 grams) and analysed by offline ESI-MS/MS. However, quite disappointingly after this G25 purification step essentially no significant m/z signals corresponding to the expected glycosylated hexapeptide could be detected unless targeted CID MS/MS scans were performed. This data indicated that the selected G-25 chromatography step was insufficient to purify the target compounds since significant amounts of co-eluting metabolites and biomolecules were present as well, suppressing ionisation and thus detection by MS analysis.

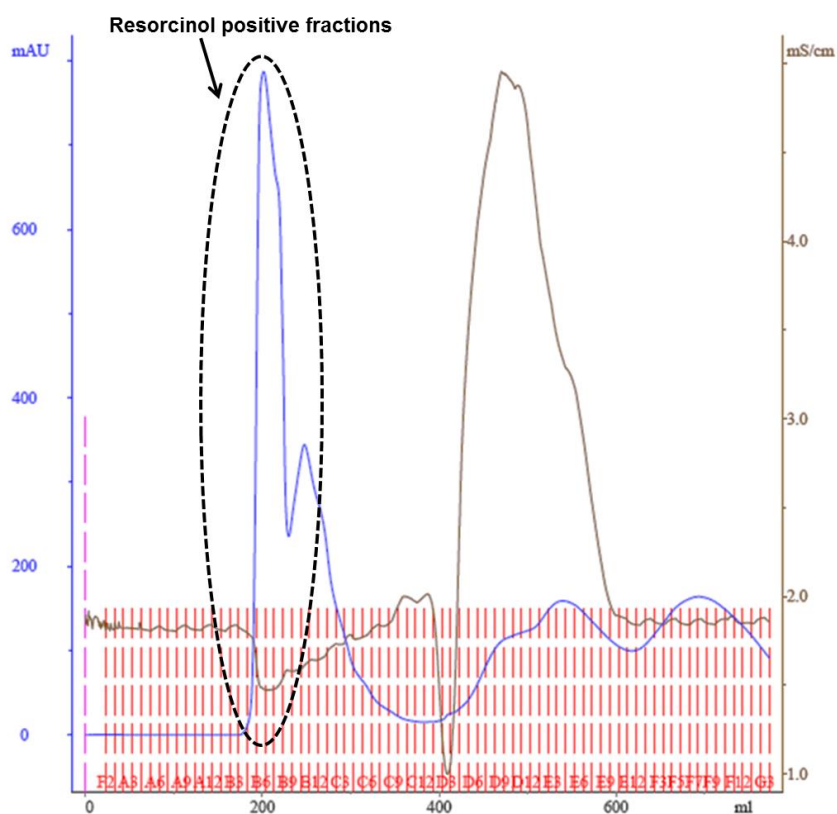


Figure 4.3: Chromatogramm of the Sephadex G-25 (fine) gel filtration of the egg yolk hexapeptide after Chloroform/methanol precipitation.

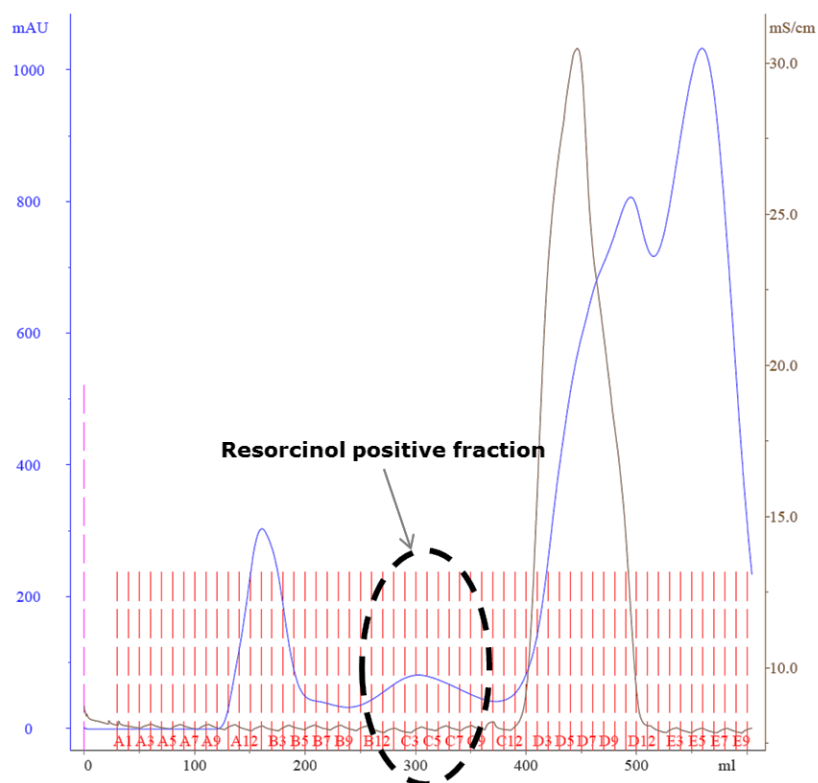


Figure 4.4: Chromatogram showing the separation of the glycosylated hexapeptide after Sephadex G-50 (fine) gel chromatography, which can be clearly separated from different larger and smaller components (peaks before and after the resorcinol positive fractions)

In order to improve the quality and yield of a purer form of glycosylated hexapeptide following the chloroform/methanol precipitation step, the same starting material was subjected to a Sephadex G-50 (Fine) gel chromatography purification (second batch). In contrast to the G-25 chromatography, three regions could be clearly distinguished in the elution profile (Figure 4.4). The resorcinol positive fraction containing the target glycopeptide was clearly separated from other contaminants. The fractions of interest were pooled and an aliquot characterised by offline ESI-MS/MS analysis (Figure 4.5), whereas the remaining sample was lyophilised. The data showed that the Wessel approach also provided access to significant amounts of the monosialylated and mono-galactosylated *N*-glycans that were also attached to the hexapeptide. These were just present in trace amounts when isolated by the Seko approach (Figure 4.2). However, since a different batch of eggs was used it cannot be excluded that this might also be a result of inter-egg variations or by influenced by storage time and temperature. The original protein source of these egg yolk hexapeptides is believed to be the glycoprotein vitellogenin II, from which by enzymatic hydrolysis the glycosylated

hexapeptide is yielded (8, 16). Thus the observed heterogeneity could potentially also be explained by the parallel action of various glycosidases. However, for the purpose of isolating glycosylated amino acids for subsequent use in solid phase glycopeptide synthesis the observed glycan structure heterogeneity did not present a limiting factor as it increases the structural diversity of the available building blocks. In addition, these truncated structures also presented easily accessible sources for future targeted enzymatic elongation to yield an even larger variety of glycosylated amino acid building blocks.

This novel isolation workflow came with the additional advantage that just a single chromatographic step after chloroform/methanol precipitation was sufficient to yield the glycosylated hexapeptide; this approach enabled the quick and easy isolation of approx. 250 mg glycopeptides from 30 egg yolks, which is in good agreement with the previously described yield of approx. 8 mg of glycopeptide/egg yolk. The obtained protein precipitate could also be further utilised for the isolation of different glycosylated amino acids, further increasing the diversity of the glycosylated amino acid building blocks available for SPGPS. In addition, one of the biggest advantages is the fact that the chloroform containing the lipids could easily be reused after distillation, thus tremendously reducing the amounts of organic solvents required for glycopeptide purification and reducing organic waste in general. The extraction of the glycosylated hexapeptide is based upon the large hydrophilicity contributed by the *N*-glycan. Upon various trials, it was observed that an optimised amount of water is necessary for the extraction of glycosylated hexapeptide in high yields. The optimised ratio for extraction of the glycosylated hexapeptide from egg yolk thus was determined to be egg yolk: water: methanol: chloroform = 1:1:2:1.

The new method for isolating the glycosylated hexapeptide was further simplified by testing the suitability of commercially available egg yolk powder. In this approach, first the dried egg yolk powder was first re-hydrated using 1:1 ratio water and stirred constantly using magnetic stirrer for 1 hr at room temperature. Following that rehydration phase, the glycosylated hexapeptide was isolated using the optimised ratio of water: methanol: chloroform mixture. This approach yielded approx. 200 mg hexapeptide from 250 g dried egg yolk powder. The suitability of

dried egg yolk power for this purpose means that also no egg white is wasted and the entire procedure is easier to perform (Table 4.1).

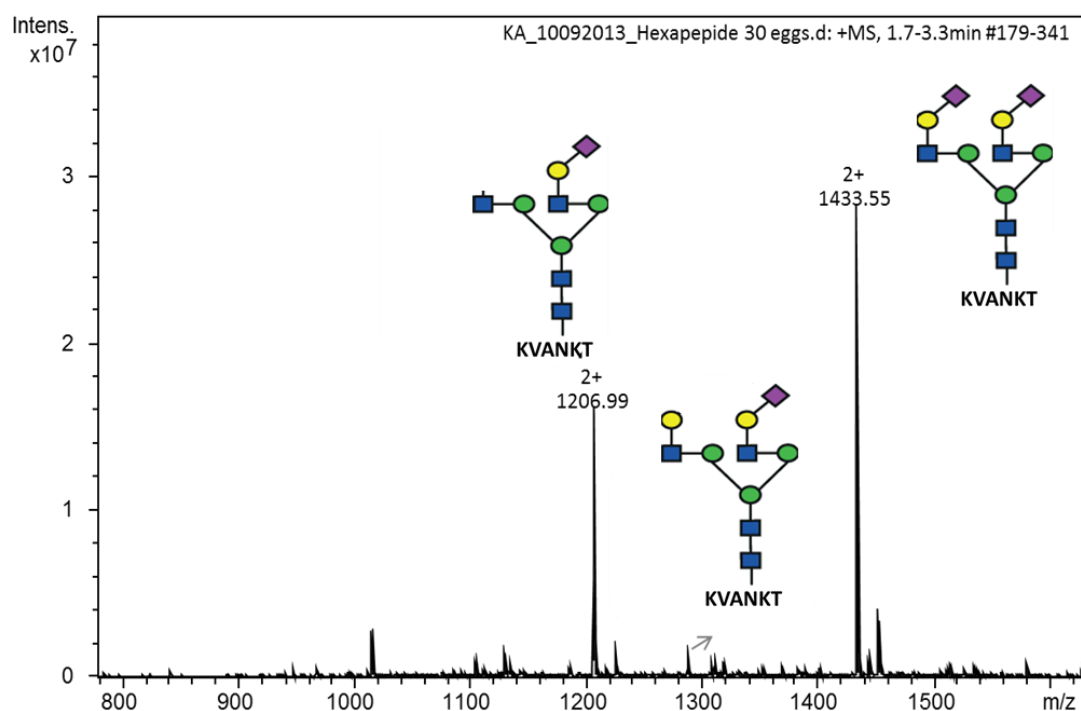


Figure 4.5: MS spectra of resorcinol positive fractions pooled after G-50 gel filtration. The signals at m/z 1433.55, 1288.55, 1206.99 $[M+2H]^{2+}$ correspond to the egg yolk hexapeptide carrying a biantennary disialylated, biantennary monosialylated and complex type mono sialylated-GlcNAc terminated N-glycan, respectively.

Table 4.1: Comparison of Seko and modified wessel approach for SGP isolation from chicken egg yolk

Parameters	Seko approach	Modified wessel approach	
		method 1	Method 2
Starting material	30 eggs	30 eggs	250 g dried egg yolk powder
Duration for extracting hexapeptide	5 days	2 days	1 day
Solvent used	phenol	chloroform/methanol	chloroform/methanol
Yield	300 mg	250 mg	200 mg

4.3.2. SYNTHESIS OF FMOC PROTECTED GLYCOSYLATED AMINO ACID BUILDING BLOCKS

To obtain the Fmoc protected glycosylated building block required for SPGPS first the isolated glycopeptide needed to be extensively proteolytically digested by Pronase to obtain just the single glycosylated Asn. The digestion reaction was monitored regularly by ESI-MS/MS, after 8 days the starting material completely disappeared yielding the desired asparagine carrying bi-antennary sialylated *N*-glycan. The reaction mixture was then further purified by Sephadex G-50 gel chromatography and the resorcinol positive fractions containing the glycosylated Asn were pooled, lyophilised (yield 51%) and the compounds checked by offline ESI-MS/MS (Figure 4.6 -4.9). These data proved that the new method provided access to considerable amounts of building blocks in much shorter time compared to pure chemical synthesis approaches (17).

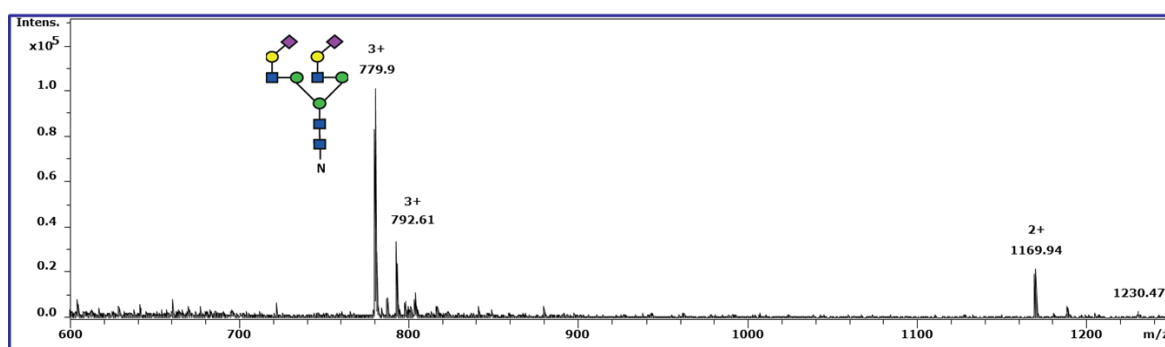


Figure 4.6: MS spectra of disialylated Asn derived from the isolated egg yolk hexapeptide (Seko approach) after extensive pronase digestion. The triply charged signal at m/z 779.28 $[M+3H]^{3+}$ corresponded to Asn containing the sialylated biantennary N-Glycan, the signal at m/z 792.28 $[M+2H+K]^{3+}$ corresponded to its potassium adduct where one proton is replaced by the potassium.

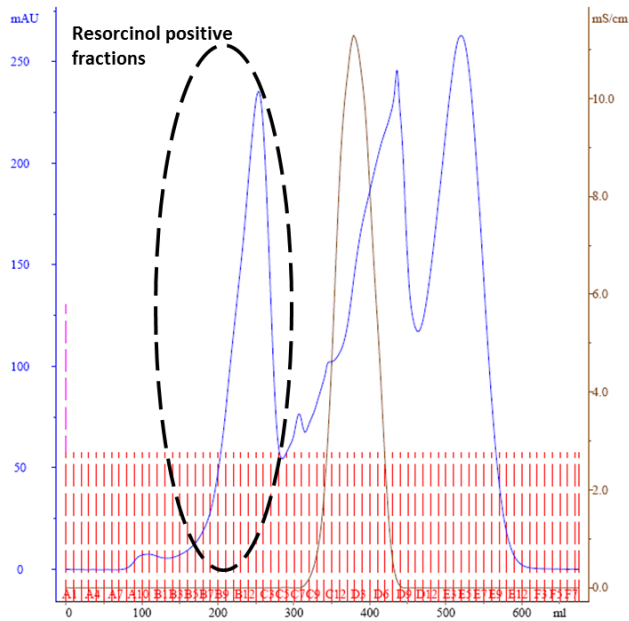


Figure 4.7: Chromatogram of the Sephadex G-50 gel chromatography showing the separation of the sample obtained after extensive digestion of the glycosylated egg yolk hexapeptide with pronase.

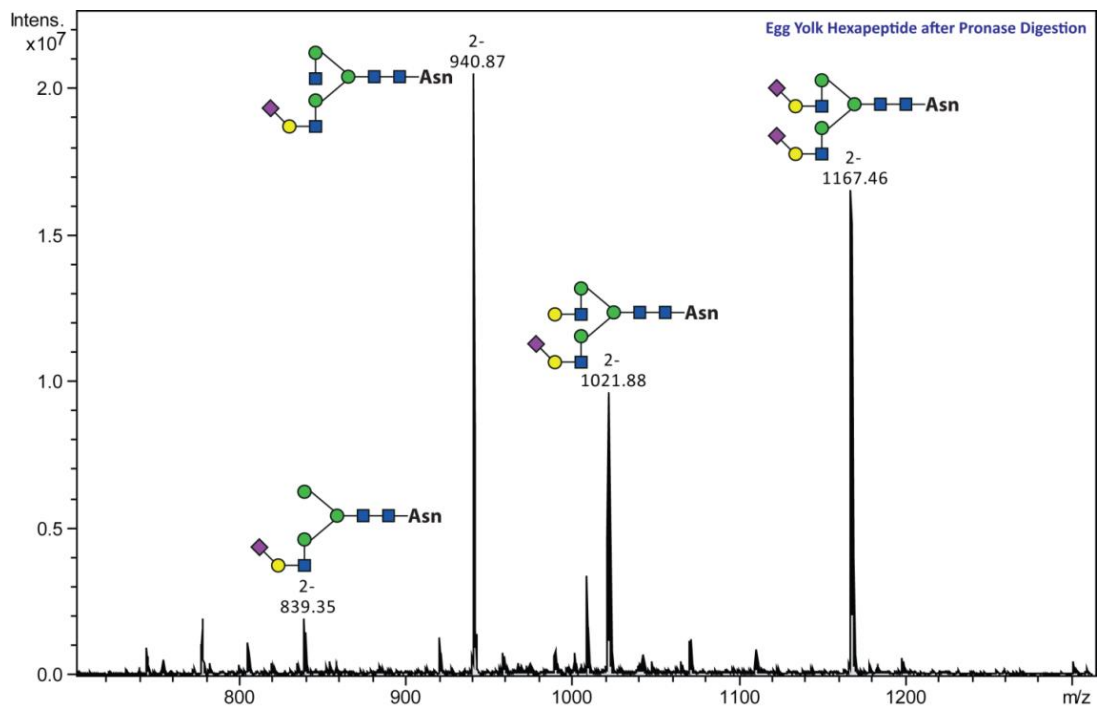


Figure 4.8: Different N-glycans attached to a single Asn amino acid that has been derived from the hexapeptide (Modified Wessel approach) after extensive pronase digestion. This analysis performed in negative mode MS. Structural assignments of the glycosylated amino acids were performed based on the observed mass and tandem mass spectra (refer to Figure 4.9).

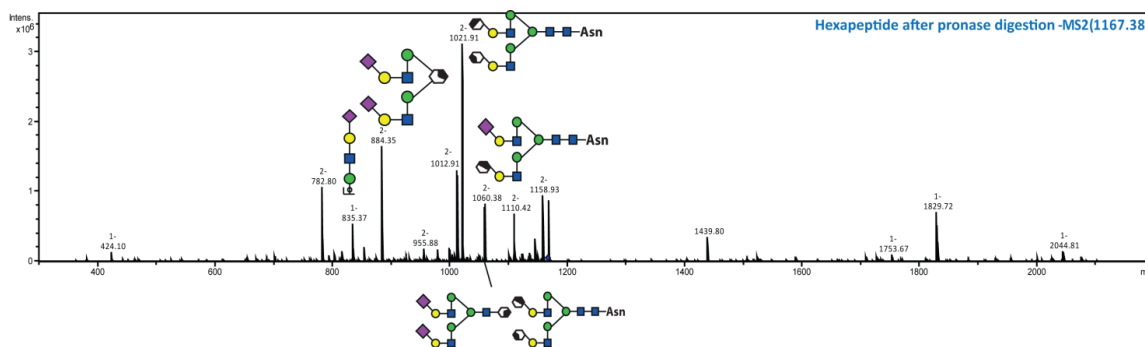


Figure 4.9: Tandem MS (MS2) spectra of m/z 1167.38 $[M+2H]^{2+}$ corresponding to the Asn carrying biantennary disialylated *N*-glycan following extensive pronase digestion of the hexapeptide enriched from egg yolk and purified as described in detail.

4.3.3. GLYCOPEPTIDE SYNTHESIS

Target glycopeptides were synthesised using 9-fluorenylmethyl chloroformate (Fmoc) based solid-phase peptide synthesis. Major steps involved in glycopeptide synthesis are (i) Isolation of the glycoamino acid building block, (ii) F-moc protection of the $N\alpha$ amino group of the glycosylated asparagine, (iii) selective benzyl esterification of sialic acids and (iv) solid phase glycopeptide synthesis (Refer section 1.5).

4.3.3.1. Fmoc protection of glycosylated amino acids

The $N\alpha$ amino group of the glycosylated asparagine was protected using 9-fluorenylmethyl-succinimidyl-carbonate (FmocOsu) and trimethyl amine in a mixture of dioxane/water (1:1) [Figure 4.10-A]. The Fmoc protected *N*-glycan Asn building blocks were purified by solid phase extraction (SPE) using 5 g C18 Sep-Pack Cartridges. The compounds were sequentially eluted using 5 and 10% acetonitrile and lyophilised to yield the Fmoc protected bi-antennary disialylated asparagine building block (80% yield). The presence of the desired compound was further verified by ESI-MS/MS analysis (Figure 4.10-B and C).

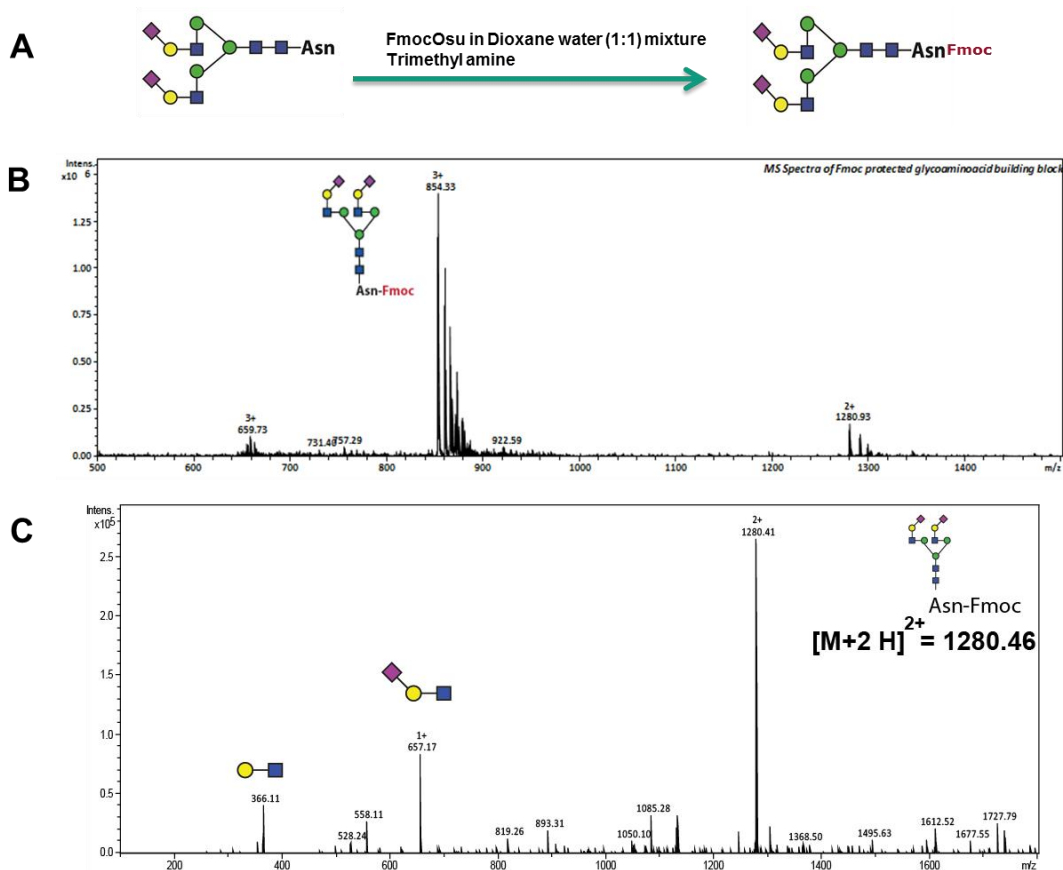


Figure 4.10: (A) Schematic representation of the Fmoc protection of Asn. (B) MS spectra of Fmoc protected Asn carrying a disialylated, biantennary N-glycan. The signals at m/z 854.02 $[M+3H]^{3+}$ and 1280.46 $[M+2H]^{2+}$ corresponded to the Fmoc protected glycosylated Asn. Other observed peaks corresponded to various combinations of sodium and potassium adducts of the target compound. (C) CID MSMS spectrum of m/z 1280.46 $[M+2H]^{2+}$ corresponding to the Fmoc protected glycosylated Asn

4.3.3.2. Esterification of sialic acid

The sialic acid residues present on the purified glycosylated amino acid building block from egg yolk were selectively protected using benzyl esterification as described by Kajihara *et al* [2]. The amino acid was dissolved in water, titrated with a solution of cesium carbonate and then lyophilized. The dry cesium salt obtained was converted to the corresponding allyl ester by reaction with allyl bromide in DMF at room temperature under argon atmosphere (Figure 4.11-A). After the completion of the reaction (24 h), the glycosyl amino acid ester was precipitated using cold diethyl ether from the reaction mixture (61% yield), purified and analyzed via ESI-MS/MS analysis (Figure 4.11-B).

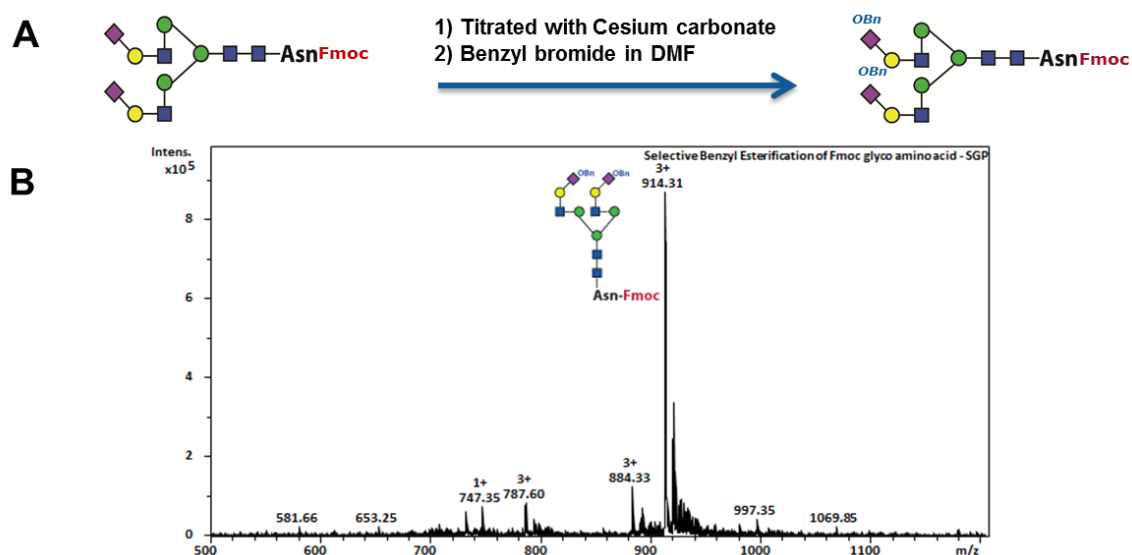
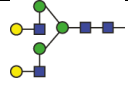

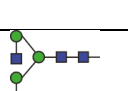
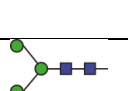
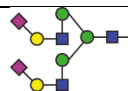

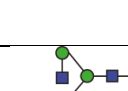
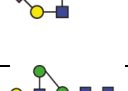
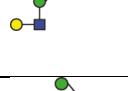

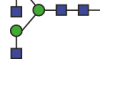


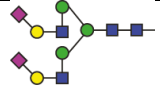
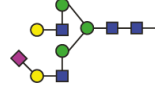
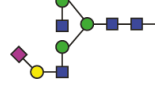
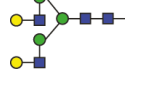

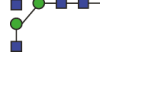
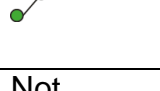
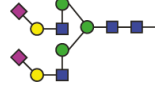
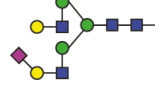
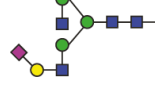
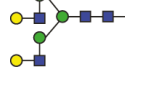


Figure 4.11: (A) Schematic representation benzyl esterification of the sialic acid (B) MS spectra of selective benzyl esterified Fmoc protected Asn carrying a disialylated, biantennary N-glycan (m/z 914.06 [$M+3H$] $^{3+}$).

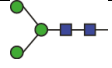
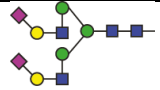
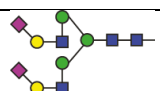
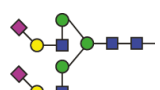
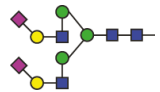
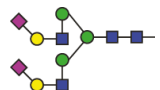

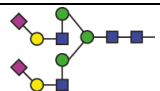

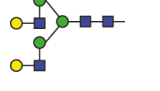
The peptide sequences were synthesised automatically using synthesiser until the point of glycosylation and further elongation of the peptide sequence including the coupling of glycosylated amino acids were performed manually. Prior the coupling of the glycosylated amino acid, the purity of the synthesised peptide was determined by ESI-MS/MS from a small amount of peptide cleaved off from the resin. A comprehensive list of peptides and glycopeptides synthesised used in the subsequent experiments is listed in the table 4.2.

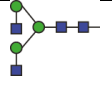
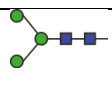

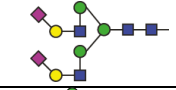
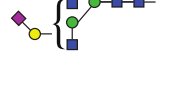
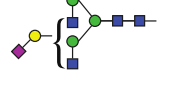
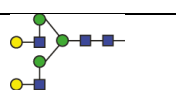
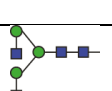
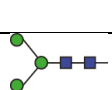
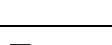
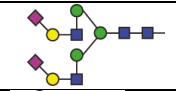

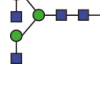
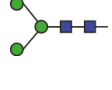

Table 1: Glycopeptides synthesised in the course of this work. A bold, underlined asparagine indicates the site of glycosylation. The N-glycan structure depicts the glycan that has been attached.

Glycopeptide	Sequence	N-Glycan	M [Da]	Protein
1	EEQY <u>N</u> STYR		3393.2771	P01857 human IgG1
2			3102.1817	
3			2940.1289	

4			2811.0863		
5			2737.0495		
6			2486.9807		
7			2080.8219		
8		Not glycosylated	1188.5047		
9	EEQY <u>D</u> STYR	Deamidated	1189.4887		
10	EEQF <u>N</u> STFR		3361.2873		P01859 human IgG2
11			3070.1919		
12			2908.1391		
13			2779.0965		
14			2705.0597		
15			2454.9909		
16			2048.8321		
17			Not glycosylated	1156.5149	
18		EEQF <u>D</u> STFR	Deamidated	1157.4989	

19	EEQY <u>N</u> STFR		3377.2822	P01860 human IgG3
20			3086.1868	
21			2924.1340	
22			2795.0914	
23			2721.0546	
24			2470.9858	
25			2064.8270	
26		Not glycosylated	1172.5098	
27	EEQY <u>D</u> STFR	Deamidated	1173.4938	
28	EEQF <u>N</u> STYR		3377.2822	P01861 human IgG4
29			3086.1868	
30			2924.1340	
31			2795.0914	
32			2721.0546	
33			2470.9858	

34			2064.8270	
35		Not glycosylated	1172.5098	
36	EEQFDSTYR	Deamidated	1173.4938	
37	ENISDPTSPLR		3432.3819	P01591 human IgJ chain
38	IIVPLNNRENISDPTSPLR		4351.9422	
39	NEEYNK		3000.1123	P02763 human alpha-1-acid glycoprotein 1
40	NISDGFDGIPDNV DAALALPAHSYSGR		4976.0874	P04004 human Vitronectin
41	NGSLFAFR		3115.2384	P04004 human Vitronectin
42	SLTFNETYQDISELVYGAK		4381.8252	P01008 human Antithrombin- III
43	NVTWK		2851.1162	P50454 Serpine H1
44*	ENYSVVFVHPK		3423.3757	artificial
45			2841.1849	

46			2517.0793	
47			2110.9205	
48			1421.6827	
49*	EVFVHP <u>N</u> YSK		3423.3757	P04070 human Protein C
50			2970.2275	
51			2970.2275	
52			2841.1849	
53			2517.0793	
54			2110.9205	
55			1421.6827	
56*				
57			2841.1849	
58	EVFVHP <u>P</u> YSNK		2517.0793	
59			2110.9205	
60			1421.6827	
<i>*Glycopeptides 44, 49 and 56 were synthesised by Mr. Hannes Hinneburg</i>				

4.4. CONCLUSION

A simple and environmentally friendlier workflow for producing sufficient amounts of glycosylated Asn building blocks carrying a disialylated, biantennary *N*-glycan was established. The availability of this building blocks allowed to initiate a comprehensive glycopeptide library that is being developed for functional and analytical glycobiology. This approach enables the specific tailoring of both the glycan and peptide moiety of a synthetic glycopeptide and thus this capacity provides an essential step forward towards studying and understanding the role of protein glycosylation.

4.4.1. SYNTHETIC GLYCOPEPTIDES ARE INDISPENSABLE GLYCOPROTEOMICS TOOLS

The generated synthetic glycopeptide library was subsequently used to systematically investigate and optimise various sample preparation steps and analytical parameters that crucially influence mass spectrometry based glycopeptide analysis. These achievements are described in detail in the chapter 5 and 6.

4.5. REFERENCE

1. Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al. *Essentials of Glycobiology*, 2nd edition. 2 ed. New York: Cold Spring Harbor Laboratory Press; 2009. 784 p.
2. Almeida A, Kolarich D. The promise of protein glycosylation for personalised medicine. *Bba-Gen Subjects*. 2016;1860(8):1583-95.
3. Deshpande N, Jensen PH, Packer NH, Kolarich D. GlycoSpectrumScan: fishing glycopeptides from MS spectra of protease digests of human colostrum slgA. *J Proteome Res*. 2010;9(2):1063-75.
4. Kolarich D, Altmann F, Sunderasan E. Structural analysis of the glycoprotein allergen Hev b 4 from natural rubber latex by mass spectrometry. *Biochim Biophys Acta*. 2006;1760(4):715-20.
5. Kolarich D, Jensen PH, Altmann F, Packer NH. Determination of site-specific glycan heterogeneity on glycoproteins. *Nat Protoc*. 2012;7(7):1285-98.
6. Kolarich D, Weber A, Pabst M, Stadlmann J, Teschner W, Ehrlich H, et al. Glycoproteomic characterization of butyrylcholinesterase from human plasma. *Proteomics*. 2008;8(2):254-63.
7. Kolarich D, Weber A, Turecek PL, Schwarz HP, Altmann F. Comprehensive glyco-proteomic analysis of human alpha1-antitrypsin and its charge isoforms. *Proteomics*. 2006;6(11):3369-80.
8. Seko A, Koketsu M, Nishizono M, Enoki Y, Ibrahim HR, Juneja LR, et al. Occurrence of a sialylglycopeptide and free sialylglycans in hen's egg yolk. *Bba-Gen Subjects*. 1997;1335(1-2):23-32.
9. Kajihara Y, Suzuki Y, Yamamoto N, Sasaki K, Sakakibara T, Juneja LR. Prompt chemoenzymatic synthesis of diverse complex-type oligosaccharides and its application to the solid-phase synthesis of a glycopeptide with asn-linked sialyl-undeca- and asialo-nonasaccharides. *Chem-Eur J*. 2004;10(4):971-85.
10. Wessel D, Flugge UI. A Method for the Quantitative Recovery of Protein in Dilute-Solution in the Presence of Detergents and Lipids. *Anal Biochem*. 1984;138(1):141-3.
11. Yamamoto N, Ohmori Y, Sakakibara T, Sasaki K, Juneja LR, Kajihara Y. Solid-phase synthesis of sialylglycopeptides through selective esterification of the sialic acid residues of an Asn-linked complex-type sialyloligosaccharide. *Angew Chem Int Edit*. 2003;42(22):2537-40.
12. Stavenhagen K, Hinneburg H, Thaysen-Andersen M, Hartmann L, Silva DV, Fuchser J, et al. Quantitative mapping of glycoprotein micro-heterogeneity and macro-heterogeneity: an evaluation of mass spectrometry signal strengths using synthetic peptides and glycopeptides. *Journal of Mass Spectrometry*. 2013;48(6):627-39.
13. Hinneburg H, Stavenhagen K, Schweiger-Hufnagel U, Pengelley S, Jabs W, Seeberger PH, et al. The Art of Destruction: Optimizing Collision Energies in Quadrupole-Time of Flight (Q-TOF) Instruments for Glycopeptide-Based Glycoproteomics. *J Am Soc Mass Spectrom*. 2016;27(3):507-19.
14. Hinneburg H, Hofmann J, Struwe WB, Thader A, Altmann F, Silva DV, et al. Distinguishing N-acetylneuraminic acid linkage isomers on glycopeptides by ion mobility-mass spectrometry. *Chem Commun*. 2016;52(23):4381-4.

15. Piontek C, Silva DV, Heinlein C, Pohner C, Mezzato S, Ring P, et al. Semisynthesis of a Homogeneous Glycoprotein Enzyme: Ribonuclease C: Part 2. *Angew Chem Int Edit.* 2009;48(11):1941-5.
16. van het Schip FD, Samallo J, Broos J, Ophuis J, Mojet M, Gruber M, et al. Nucleotide sequence of a chicken vitellogenin gene and derived amino acid sequence of the encoded yolk precursor protein. *J Mol Biol.* 1987;196(2):245-60.
17. Wang P, Zhu J, Yuan Y, Danishefsky SJ. Total synthesis of the 2,6-sialylated immunoglobulin G glycopeptide fragment in homogeneous form. *J Am Chem Soc.* 2009;131(46):16669-71.

5. IT'S ALL ABOUT THE SOLVENT: ON THE IMPORTANCE OF THE MOBILE PHASE FOR ZIC-HILIC GLYCOPEPTIDE ENRICHMENT

This chapter has been modified in part from the following article:

Alagesan K, Khilji SK, and Kolarich D. (2016). It is all about the solvent: on the importance of mobile phase for ZIC-HILIC glycopeptide enrichment. - Anal Bioanal Chem, DOI : 10.1007/s00216-016-0051-6 – *Accepted Manuscript*

5.1. INTRODUCTION

Peptide and glycopeptide mixtures are frequently analysed following proteolytic digestion by either LC-ESI-MS/MS or MALDI-TOF-MS. However, simultaneous detection of peptides and glycopeptides can be tricky. Glycoprotein proteolysis often results in unequal mixtures of these compounds as glycopeptide microheterogeneity reduces the concentration of each individual glycopeptide molecule compared to unmodified peptides obtained by the same digest (1). Hydrophobic molecules also tend to provide stronger signals compared to hydrophilic ones, which further complicates glycopeptide detection in the presence of unmodified peptides (2). Subsequently, glycopeptide signal strengths are significantly lower compared to their unmodified counterparts, mostly due to the presence of the large hydrophilic glycan moiety (3). Therefore glycopeptide enrichment is often performed to allow their detection and identification (4-6).

In contrast to glycopeptide enrichment methods using lectins, hydrazide chemistry, titanium dioxide or graphitised carbon, Hydrophilic Interaction Chromatography (HILIC) comes with the unique advantage to enable glycopeptide enrichment in a largely glycan structure unbiased manner. During the HILIC enrichment process glycopeptides are also not chemically or enzymatically altered - this is highly relevant for in-depth glycoproteomics. Another significant advantage of HILIC is that both peptide and glycan present in the enriched fraction can be analysed in a high throughput fashion as intact glycopeptides but also individually after

enzymatic treatments with PNGase F/A. In contrast to normal phase chromatography, the HILIC retention mechanism is largely a result of a hydrophilic partitioning of the analyte to the water enriched layer surrounding the polar stationary phase (7). Glycopeptide retention mainly depends on the size of the glycan moiety and its hydrophilic properties, but also the hydrophilic features of the peptide backbone. The polar interaction between the glycopeptides' glycan moieties with the hydrophilic layer surrounding the stationary phase provides an opportunity to separate glycopeptides from (usually) less hydrophilic peptides. Furthermore, depending upon the type of the used stationary phase, hydrogen bonding, electrostatic or dipole-dipole interactions also influence analyte retention (7, 8). A wide range of HILIC stationary phases have successfully been reported for glycopeptide enrichment ranging from silica particles (9), cellulose (10), sulfoalkylbetaine (ZIC-HILIC) (11, 12), amide-based (13) to even simple cotton (14). Compared to other HILIC materials, ZIC-HILC is known to provide higher selectivity for glycopeptides (15).

Glycopeptide enrichment is usually performed at starting conditions with 80% organic solvent concentration while elution is performed by disturbing the hydrophilic interactions using aqueous conditions. Acetonitrile is by far the most popular organic mobile phase applied for this purpose (16). Although HILIC SPE is efficient (reproducible and sensitive) for glycopeptide enrichment from mixtures, hydrophilic non-glycosylated peptides are also frequently co-enriched. This represents a particular problem for analysing complex samples as the co-enriched hydrophilic peptides can cause glycopeptide ion-suppression. Co-enrichment of hydrophilic non-glycosylated peptides can, however, be avoided or at least significantly reduced by the addition of suitable ion-pairing reagents such as trifluoroacetic acid (TFA) or hydrochloric acid (HCl) (17, 18). Alternatively, the reduction of non-specific enrichment has also been reported by digesting the glycoprotein with non-specific proteases prior HILIC enrichment. This greatly increases glycopeptide hydrophilicity due to the shorter peptide backbone (19). The drawbacks of this approach are, however, increased sample heterogeneity, impeded accurate site specific glycan structure assignment and lack of accurate relative quantitation of site specific microheterogeneity (20).

Andrew Alpert once described HILIC as "the combination of hydrophilic stationary phases and hydrophobic, mostly organic mobile phases"(7). However, a systematic evaluation of mobile phase effect on glycopeptide enrichment selectivity and efficiency is lacking. In this work a systematic evaluation of various mobile phases [Acetonitrile (ACN), Methanol (MeOH), Ethanol (EtOH) and Isopropanol (IPA)] effects on glycopeptide enrichment selectivity and efficiency using ZIC-HILIC is presented. Glycopeptide enrichment efficiencies were evaluated for each solvent system using a variety of samples, which required the development of an enrichment technique suitable for this purpose termed "Drop-HILIC". Drop-HILIC is significantly cheaper and quicker to perform than the conventional micro-spin technique and provides comparable results. Different purified glycoproteins as well as more complex samples provided in the form of depleted and non-depleted human serum were tested to conclude that glycopeptide enrichment efficiency largely depends on the organic mobile phase.

5.2. MATERIALS AND METHODS

5.2.1. MATERIALS

If not otherwise stated, all materials were purchased in the highest possible quality from Sigma-Aldrich (St. Louis, MO, USA). Trypsin (sequencing grade) was obtained from Roche Diagnostic GmbH (Mannheim, Germany). Water was used after purification with a Milli Q-8 direct system (Merck KGaA, Darmstadt, Germany). IgG (I4506) and A1PI (A6150) were obtained from Sigma-Aldrich. Human serum was obtained from BioreclamationIVT (New York, USA). The amino acid numbering applied for all proteins analysed in this study is based on the respective UniProtKB entries.

5.2.2. HIGH ABUNDANCE SERUM PROTEIN DEPLETION

Depletion of abundant serum proteins was performed according to the manufactures instructions using a commercially available kit (ProteoSpin™ Abundant Serum Protein Depletion Kit Cat. # 17300).

5.2.3. IN-SOLUTION PROTEASE DIGESTION

One hundred microgram of protein [IgG, A1PI or serum (depleted and non-depleted)] were reduced with 10 µL of 500 mM dithiothreitol (DTT) (in H₂O) (99°C, 5 min) and then subsequently alkylated with 10 µL 500 mM iodoacetamide (IAA)

solution (in H₂O) at room temperature for 60 min in dark. Prior trypsin digestion, the samples were subjected to chloroform-methanol precipitation as described earlier (21). The protein pellet was resolubilised in 200 µL of 25 mM NH₄HCO₃ and trypsin added in a 1:30 ratio (enzyme:substrate). After overnight incubation at 37°C the resulting glycopeptide/peptide mixtures were aliquoted corresponding to 3 µg of initial protein concentration and dried in the speedvac without additional heating. The samples were stored at -25°C until further experiments.

5.2.4. HILIC ENRICHMENT – MICRO SPIN

ZIC-HILIC (pore size 200 Å, 10 µm particle size, SeQuant AB, Sweden) was filled up to 1.5 cm in a C18 ZipTip P10 (Merck Millipore, Tullagreen, IRL). The column was washed three times with 50 µL of 1% TFA and then equilibrated three times with 50 µL of 80% ACN containing 1% TFA. The dried sample was reconstituted in 10 µL 1% TFA and slowly adjusted to 80% ACN/1% TFA by the addition of 40 µL ACN/1% TFA. The sample was applied onto the column and centrifuged until the entire liquid passed through. The flowthrough was reapplied onto the column and again centrifuged. The sample was washed twice with 50 µL of 80% ACN containing 1% TFA and glycopeptides were eluted off the column by washing it thrice with 50 µL of 1% TFA followed by 50 µL of 80% ACN containing 1% TFA. The eluted fraction was dried in the speedvac and reconstituted in 50 µL of 0.1% TFA for further MS analyses. All analyses were performed in triplicate.

5.2.5. HILIC ENRICHMENT – "DROP-HILIC"

Tryptic protein digests were dissolved in 10 µL 1% TFA and slowly adjusted to 80% organic solvent conditions by the addition of 40 µL organic solvent (ACN/1% TFA or EtOH/1% TFA or MeOH/1% TFA or IPA/1% TFA). ZIC-HILIC beads were washed three times with 1% TFA (3x 250 µL) and then equilibrated three times with appropriate binding solution (3x 250 µL). Subsequently, the HILIC beads were added to the sample and incubated at room temperature for 1, 3 or 5 min with occasional shaking. After incubation, the HILIC beads were spun down in a table centrifuge. The supernatants (flowthrough) were transferred into a new vial. The HILIC beads were then mixed with 50 µL appropriate binding solution, vortexed and spun down. The supernatants were pooled together and the washing step was repeated twice. Enriched glycopeptides were eluted using 3x 50 µL of the elution buffer (1% TFA) and all three elution supernatants were combined in a

new vial. The eluate was dried in the speedvac and reconstituted in 50 μ L of 0.1% TFA for further MS analyses. All analyses were performed in triplicate.

5.2.6. GLYCOPEPTIDE SYNTHESIS

Solid Phase Glycopeptide Synthesis (SPGPS) was performed manually using 5-mL and 10-mL disposable polypropylene syringes with a bottom filter. All peptides and glycopeptides were synthesised by SPGPS using previously reported fluorenylmethoxycarbonyl (Fmoc) protocols (3, 22, 23) and as described in detail in the chapter 4.

5.2.7. LC-MS ANALYSIS PARAMETERS

Nano-LC-ESI-MS analysis was carried out on an Ultimate 3000 RSLC-nano system (Dionex/Thermo Scientific, Sunnyvale, CA) coupled to an amaZon speed ETD ion trap mass spectrometer (IT-MS) equipped with CaptiveSpray nanoBooster™ (both Bruker, Bremen, Germany). In each run glycopeptides corresponding to 180 ng were injected.

In nano-LC–mode the peptides were concentrated on a C18 precolumn (Acclaim PepMap100™, Thermo, 100 μ m x 20 mm, 5 μ m particle size) and separated by reversed phase chromatography on a C18 analytical column (Acclaim PepMap™, Thermo, 75 μ m x 15 cm, 3 μ m particle size). The samples were loaded in 99% loading buffer (0.1% TFA) for 5 min on the precolumn at a flow rate of 5 μ L/min before the captured peptides were subjected to reversed phase nanoLC at a flowrate of 400 nL/min on a column equilibrated in 95% buffer A (0.1% formic acid). The gradient conditions were as follows: increase of buffer B (90% acetonitrile containing 0.1% formic acid) from 5% to 45% (6-36 min), further increase to 70% B (36-38 min), followed by a steeper increase to 90% B (40-42 min). The column was held at 90% B for 10 min (42-52 min). The mass spectrometer was set-up to perform CID on the three most intense signals in every MS scan. An m/z range from 400-1600 Da was used for data dependent precursor scanning. The MS data was recorded using the instrument's "enhanced resolution mode". MS/MS data was acquired in "ultra-mode" over an m/z range from 100-2000. A detailed parameter setting is provided in the Table 5.1 following MIRAGE (24) and MIPAE (25) recommendations.

Data analysis was performed using ProteinScape 3 (Bruker Daltonics) and MASCOT 2.3 (MatrixScience, United Kingdom) using the following search parameters: Cysteine as carbamidomethyl was set as fixed modification, and deamidation (Asn/Gln) and oxidation (Met) were set as variable modifications. Up to two missed cleavages were allowed. Peptide tolerance was set at ± 0.3 Da for MS and at ± 0.5 Da for MS/MS. The data were searched against the SwissProt protein database (taxonomy restriction: *Homo sapiens*, SwissProt 2011_08; 531,473 sequences; 188,463,640 residues).

Table 5.1: Ion trap settings applied in this study, following the MIRAGE guidelines (www.beilstein-mirage.org)

MS Settings - General	
ESI probe	CaptiveSpray™
Capillary voltage	1.3 kV
SPS	<i>m/z</i> 900
Compound stability	100%
Trap Drive Level	100%
Spectra averaging	5
Dry gas temperature	150°C
Dry Gas flow	3 L/min
Maximum accumulation time	200 ms
Ion mode	positive
MS-Scan	
MS Scan mode	Enhanced scan
ICC target	200000
Mass detection range	<i>m/z</i> 350-1800
MS2	
MS scan mode	ultrascan
SPS MS(n)	automatic
MS(n) spectra averages	5
MS(n) ICC target	40000
Preferred charge state	\geq Doubly,
Active exclusion	Off
Mass detection range	<i>m/z</i> 100-2000
Isolation width	3 Da
Exclude singly charged ions	on
No. of precursor ions	3
SmartFrag	Enhanced
SmartFrag Start amplitude	30%
SmartFrag End amplitude	120%
Fragmentation width	5 <i>m/z</i>

5.3. RESULTS AND DISCUSSION

5.3.1. RATIONALE AND EXPERIMENT DESIGN

The efficiency and selectivity of glycopeptide enrichment by ZIC-HILIC SPE was first investigated using well-defined synthetic glycopeptides spiked into different concentrations of tryptic peptides derived from BSA. Following these initial experiments HILIC was performed on proteolytic digests of standard glycoproteins (IgG and A1PI) and finally applied on a complex sample derived from human serum (Figure 5.1). The influence of the mobile phase on glycopeptide enrichment efficiency was tested using four different solvents. (i) Acetonitrile, (ii) Methanol, (iii) Ethanol and (iv) Isopropanol were used at a concentration of 80% in water containing 1% TFA as an acidic modifier. In the course of this work a simplified HILIC enrichment technique ("Drop-HILIC") was developed and compared it to the traditional micro-spin HILIC approach.

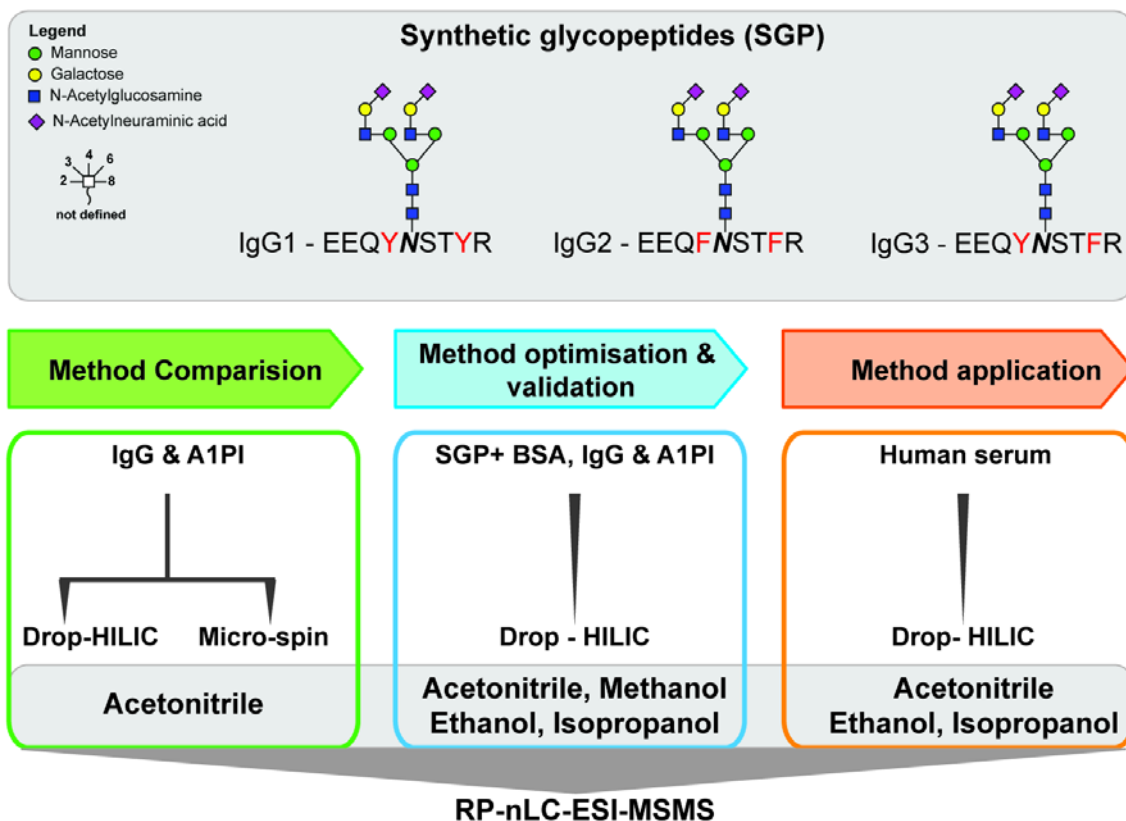


Figure 5.1: Experiment design – Investigating the influence of various organic mobile phases on ZIC-HILIC glycopeptide enrichment efficiency and development of a simplified HILIC enrichment technique ("Drop-HILIC"). All four mobile phases [(i) Acetonitrile (ii) Ethanol (iii) Methanol and (iv) Isopropanol] were tested using 80% organic solvent starting conditions. Glycopeptide enrichment efficiencies were evaluated by nanoLC ESI-MS/MS using samples of increasing complexity ranging from well-defined synthetic glycopeptides spiked into a tryptic digest of BSA over individual standard glycoproteins to a tryptic digest of depleted and non-depleted human serum.

5.3.2. SALT REMOVAL IS CRUCIAL FOR EFFICIENT HILIC ENRICHMENT

Initial experiments applying the micro-spin method to enrich IgG glycopeptides resulted in no/insufficient enrichment when performed subsequently following in-solution trypsin digestion (data not shown). Applying this workflow directly on in-gel digested (glyco)peptides, however, provided the expected results. This might be due to the presence residual of higher salt concentrations derived from the alkylating reagents resulting in electrostatic (ionic) interactions that were compromising glycopeptide enrichment when performed subsequently after an in-solution proteolytic digest. A simple reversed phase based desalting step could be introduced prior HILIC. This step, however, can also lead to the loss of very hydrophilic glycopeptides (26). To avoid such losses a modified sample

preparation protocol by introducing a simple chloroform-methanol precipitation step prior proteolysis to remove any excess DTT and IAA. This simple, fast and efficient step enabled successfully enrichment of glycopeptides, greatly minimising the sample losses and eliminate any molecules affecting the enrichment efficiency. These optimised proteolytic sample preparation conditions were then applied for all following experiments.

5.3.3. SAME, SAME BUT DIFFERENT: TO SPIN OR TO "DROP-HILIC"

Use of ACN as loading solvent approximately requires 3-5 min to pass through the ZIC-HILIC micro spin column. However, when applying more viscous solvents such as isopropanol up to 20 min were required. As the intended aim was to evaluate the influence of various organic mobile phases on the glycopeptide enrichment, a simplified and accelerated technique by simply co-incubating the sample with the HILIC beads was established (known as "Drop-HILIC"). This approach provided the opportunity to normalise incubation times and evaluate any solvent effect, which could not reasonably be achieved by the micro spin HILIC method. Besides being easier and quicker to perform, Drop-HILIC provided additional advantages: (i) the amount of HILIC material could be optimised to accommodate variable sample amounts and (ii) it was more cost and time effective as no C18 Zip-tip columns or custom made tips were required.

First, traditional micro-spin and Drop-HILIC approaches were compared using two well-characterised standard glycoproteins, human IgG and A1PI using 80% acetonitrile as loading solvent. The glycopeptide enriched fractions were subsequently analysed by nanoLC-ESI-MSMS (Figure 5.2; please refer to Appendix Table 2.1 and 2.5). Both techniques yielded comparable IgG glycopeptide profiles (Figure 5.2-A and Appendix Table 2.2). IgG2 provided the most abundant signals followed by the glycopeptides derived from IgG1 and IgG4. However, a slightly different trend was observed in the case of A1PI. The drop approach enriched glycopeptides A1PI-GP3 (²⁶⁸YLGNAIAIFFLPDEGK²⁸³) significantly better. On the other hand, glycopeptides containing a larger peptide backbone (A1PI-GP4 carrying H5N4F0Na2 and H3N3F0Na1 as well as A1PI-GP1) were not enriched with similar efficiency (Figure 5.2-B and Appendix Table 2.6). In order to evaluate whether the one-minute incubation time affects Drop-HILIC enrichment efficiency for these larger glycopeptides, possibly due to

inadequate phase partitioning, additional experiments with various incubation times were performed.

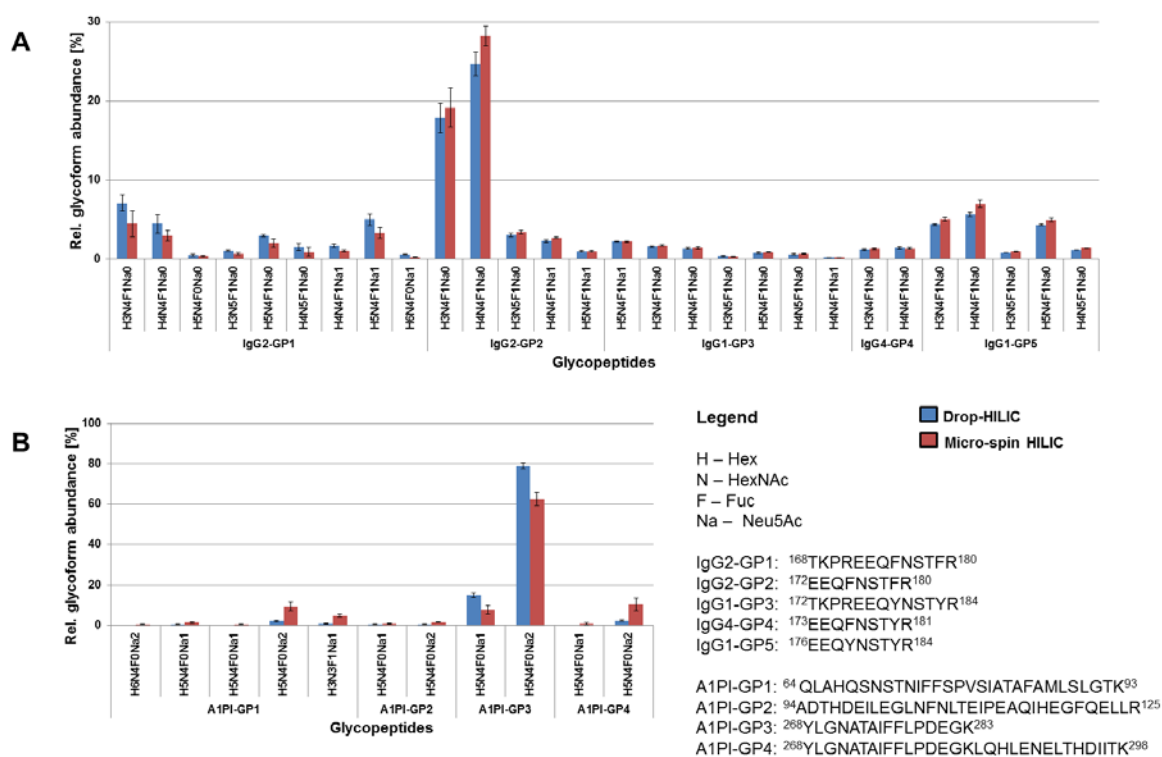


Figure 5.2: Glycoprofile comparison of the Drop-HILIC and micro-spin techniques using ACN as mobile phase for glycopeptide enrichment from (A) IgG and (B) A1PI. Enriched glycopeptides were analyzed by RP-nanoLC-ESI-MS/MS. IgG contains a single site of glycosylation whereas three are present in A1PI. The relative abundances were determined using the area under the curve of extracted ion chromatograms (EIC's) produced from all glycoform and charge state signals detected for each single glycopeptide. Three technical replicates were performed. Both techniques performed similar on IgG glycopeptides, while some glycoprofile differences were found for the hydrophobic A1PI glycopeptides.

5.3.4. INCUBATION TIME DOES NOT INFLUENCE DROP-HILIC ENRICHMENT EFFICIENCY

Various incubation times (1, 3 and 5 min) did not show any significant changes in the enrichment efficiency of IgG glycopeptides under standard conditions (Figure 5.3-A and Appendix Table 2.3). Longer incubation times did also not improve the enrichment of the larger hydrophobic A1PI glycopeptides, indicating that the Drop-HILIC approach exhibited some limitations for efficiently enriching such large, >25 amino acid long, comparably hydrophobic glycopeptides. Despite these observed limitations the Drop-HILIC approach provided glyco-profile results comparable to

the classical micro-spin HILIC method for IgG and A1PI-GP3 (Figure 5.3-B and Appendix Table 2.7). Drop-HILIC, however, came with the advantages of being significantly quicker and cheaper to perform. As the incubation time did not show any influence on the enrichment efficiency all further experiments were performed using 1 min incubation times for the evaluation of any effect the organic mobile phase has on glycopeptide enrichment efficiency.

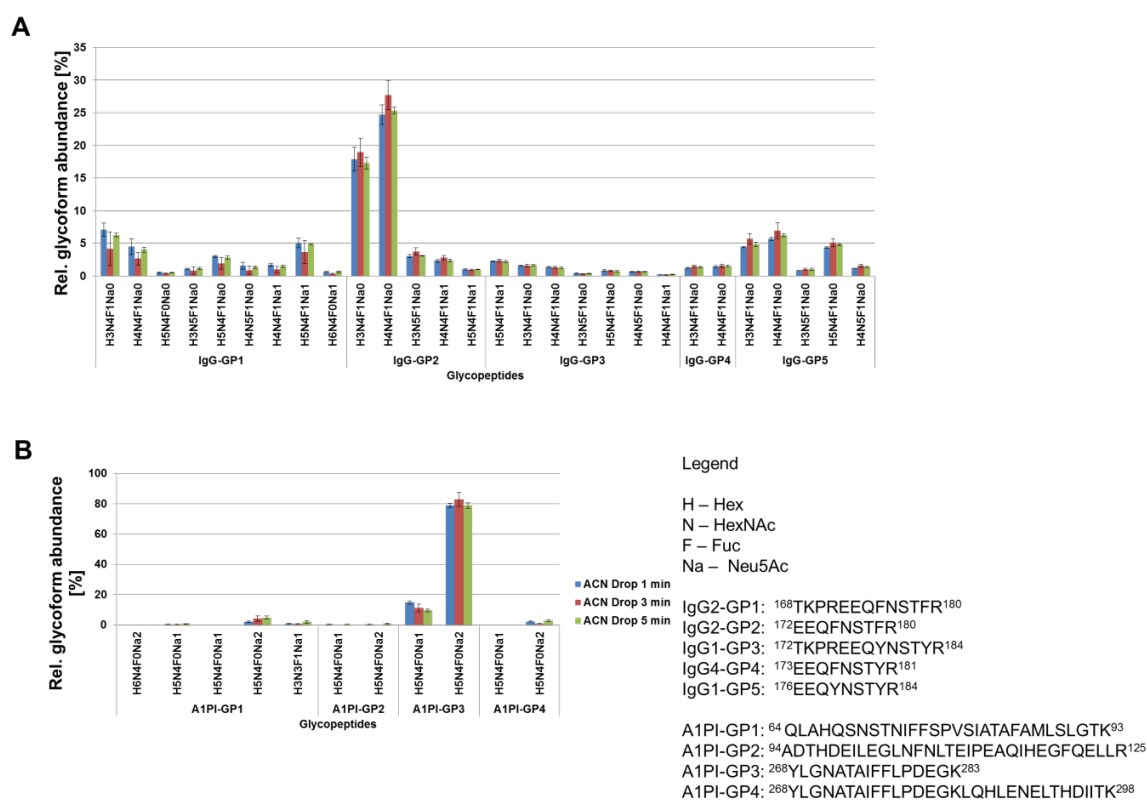


Figure 5.3: Effect of incubation on glycopeptide enrichment efficiency using Drop HILIC. Variation in incubation time did not improve the glycopeptide enrichment efficiency using Drop-HILIC approach for various glycopeptide species derived from (A) IgG and (B) A1PI samples

5.3.5. IT'S ALL ABOUT THE SOLVENT – INFLUENCE OF THE SOLVENT SYSTEM ON ZIC-HILIC GLYCOPEPTIDE ENRICHMENT

The underlying mechanism of analyte retention in ZIC-HILIC is originating from hydrophilic partitioning in addition to contributions derived from minor electrostatic interactions. The use of TFA in the mobile phase nullifies any possible electrostatic interactions making hydrophilic partitioning the only cause for analyte retention (18). For an efficient enrichment the ideal mobile phase should be water miscible but not contribute any hydrogen donor or acceptor functionalities. A completely

aprotic solvent such as acetonitrile embraces this particular feature and thus is often used for glycopeptide enrichment by HILIC SPE. However, a systematic evaluation of any mobile phase effect on glycopeptide enrichment is still lacking. The optimised sample preparation step in combination with simple, fast & equally efficient Drop-HILIC approach enabled the evaluation of the influence of various MS compatible mobile phase solvents on ZIC-HILIC glycopeptide enrichment efficiency.

First, the solvent effect was evaluated using a well-defined system containing synthetic *N*-glycopeptides corresponding to the tryptic glycopeptide sequences present in IgG1, IgG2 and IgG3 carrying a biantennary, disialylated *N*-glycan spiked into the background of tryptically digested BSA. The synthetic glycopeptides were mixed with tryptic BSA derived peptides in molar ratios 3:1:1:3 (IgG1:IgG2:IgG3:BSA) and glycopeptide enrichment was performed using the Drop-HILIC technique. Initial results indicated that methanol is a non-favoured mobile phase for this purpose (data not shown). Due to its strong tendencies to form hydrogen bonds methanol effectively competes for the active sites on the stationary phase and is thereby perturbing hydrophilic partitioning. This subsequently resulted in strongly reduced glycopeptide retention and thus was not further evaluated (27). The glycopeptide enriched fractions from the other solvents (ACN, EtOH, IPA), however, were analysed by RP-nano LC-ESI-IT-MSMS.

LC-ESI-MS analysis indicated a strong mobile phase solvent-dependency in the selectivity and efficiency for glycopeptide enrichment (Figure 5.4-A), which was also considerably influenced by the hydrophilicity of the peptide backbone (Table 5.2). The synthetic IgG1 glycopeptide was efficiently enriched in a similar manner by all three solvents, while a strong mobile phase dependency was observed for IgG2 and IgG3 synthetic glycopeptides. Ethanol significantly enriched the synthetic IgG3 (75.43% \pm 8.42) and IgG2 (88.58% \pm 6.76) glycopeptides better than ACN or IPA while the IgG2 glycopeptide could not be enriched at all using IPA in the background of tryptic BSA peptides (Figure 5.3-A). Interestingly, this peptide is also the most hydrophobic of the three synthetic *N*-glycopeptides (GRAVY score of -1.60, see Table 5.2). The data suggested that either the hydrophilic BSA peptides outcompeted the IgG2 synthetic *N*-glycopeptide or suppressed its ionisation, making it not detectable under the conditions used. When excess molar

ratios of BSA were applied (IgG1:IgG2:IgG3:BSA = 3:1:1:6 and 3:1:1:10), glycopeptide enrichment efficiency was compromised especially in the case of EtOH and IPA as the abundances of co-enriched peptides clearly increased (Appendix Figure 2.1). Under the tested conditions ACN provided the best compromise for the retention of all three synthetic glycopeptides while keeping the number of co-enriched BSA peptides at a low level. Nevertheless, with all three mobile phases numerous peptides were co-enriched in a mobile phase dependent manner (Figure 5.4-B).

Inspired from these results, the loading solvent influence on ZIC-HILIC glycopeptide enrichment using individual standard glycoprotein digests of IgG and A1PI was evaluated. Also for these low-complex samples loading solvent dependant selectivity was found. The human IgG glycopeptides showed similar results as obtained for the synthetic ones, with the exception that IPA was a suitable solvent to enrich the IgG2 glycopeptide (Figure 5.4-C and appendix Table 2.4), indicating that the BSA tryptic peptide background was interfering with its enrichment in earlier experiments (Figure 5.3-A). In the case of A1PI, however, isopropanol provided the best compromise for the simultaneous enrichment of hydrophilic and hydrophobic glycopeptides (Figure 5.4-D and Appendix Table 2.8). This can possibly be explained by the fact that not all (glyco)peptides were equally soluble under 80% organic mobile phase conditions (28). This hypothesis was also supported by the different identified co-enriched unmodified peptides that were found for the individual solvents (Table 5.3 and Figure 5.5). As a consequence of this insolubility an insufficient enrichment of certain glycopeptide species was observed when EtOH was used. These results indicate that analyte retention and subsequent glycopeptide enrichment efficiency in HILIC is just not controlled by hydrophilic partition but much more complex mechanism occurring at the interface of the stationary polar and organic mobile phase during sample solvation prior sample loading.

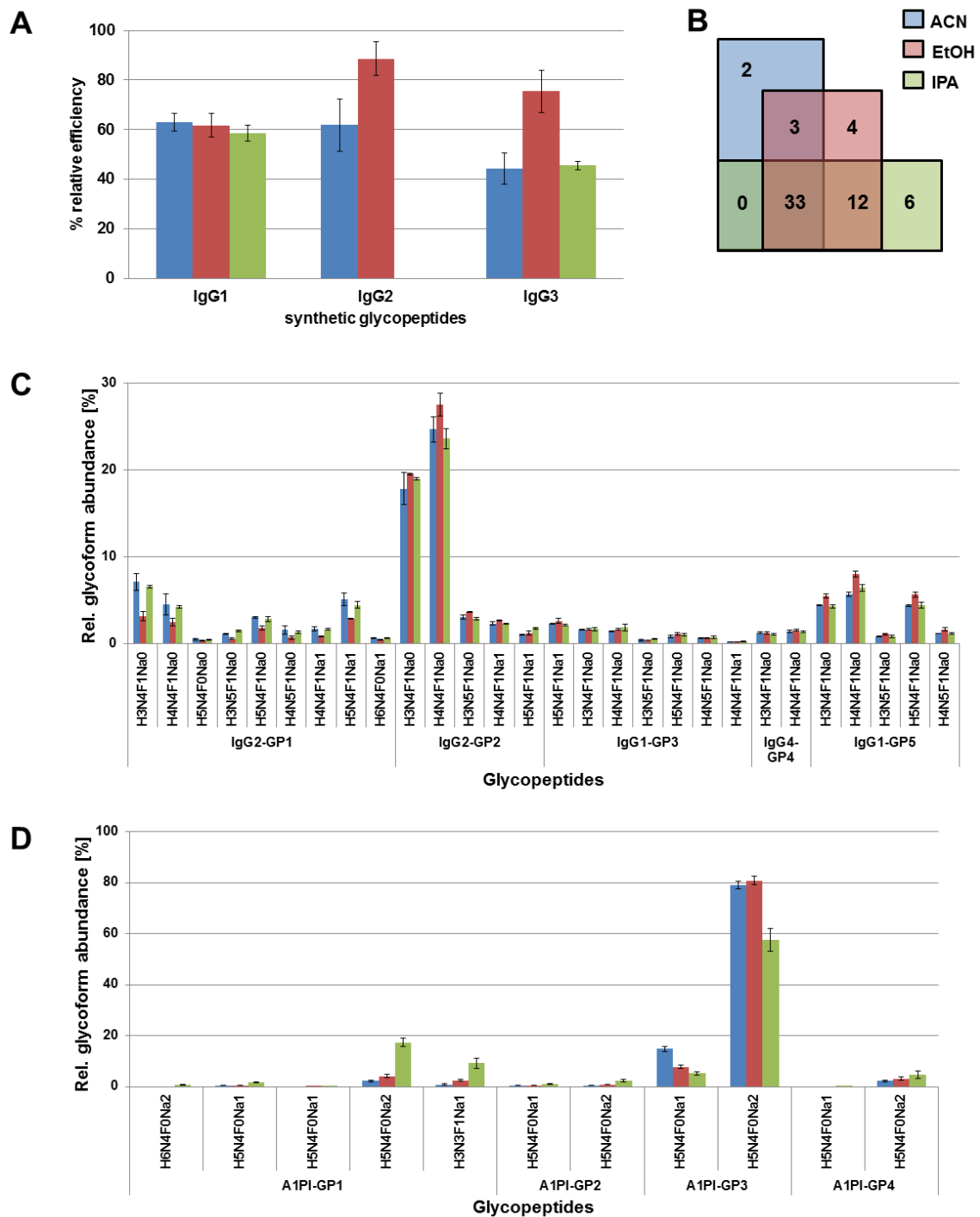


Figure 5.4: Loading solvent effect on ZIC-HILIC glycopeptide enrichment. A: Synthetic N-glycopeptides corresponding to IgG1, IgG2 and IgG3 tryptic peptides were spiked into a tryptic digest of BSA. Whereas ACN did effectively enrich all three isoforms, EtOH and IPA exhibited IgG subclass specific tendencies. B: Venn diagram showing the number of BSA derived peptides co-enriched with the synthetic glycopeptides when using different loading solvents. C+D: Different mobile phases were used for loading the (glyco)peptide mixtures onto the resin, resulting in a differential enrichment of various glycopeptide species from (C) IgG and (D) A1PI samples.

Table 5.2: GRAVY scores of non-glycosylated peptides identified by RP-nLC

Protein	Sequence NO	Sequence	Gravy Score
A1AT	64-93	QLAHQSNSTNIFFSPVSIATAFAMLSLGTK	0.43
A1AT	94-125	ADTHDEILEGLNFNLTETPEAQIHEGFQELLR	-0.47
A1AT	268-283	YLGNATAIFFLPDEGK	0.16
A1AT	268-298	YLGNATAIFFLPDEGKQLQHLENELTHDIITK	-0.20
IgG2	168-180	TKPREEQFNSTFR	-1.93
IgG2	172-180	EEQFNSTFR	-1.60
IgG1	172-184	TKPREEQYNSTYR	-2.56
IgG1	176-184	EEQYNSTYR	-2.51
IgG4	173-181	EEQFNSTYR	-2.06

Table 5.3: Number of identified peptides in the glycopeptide enriched fraction as reported by MASCOT/ProteinScape search for IgG and A1PI

Solvent	IgG			A1PI		
	EtOH	ISO	ACN	EtOH	ISO	ACN
Trial 1	35	40	41	63	72	57
Trial 2	25	45	36	58	69	49
Trial 3	31	47	40	54	68	58
Mean	30.33	44	39	58.33	69.67	54.67
SD	5.03	3.61	2.65	4.51	2.08	4.93

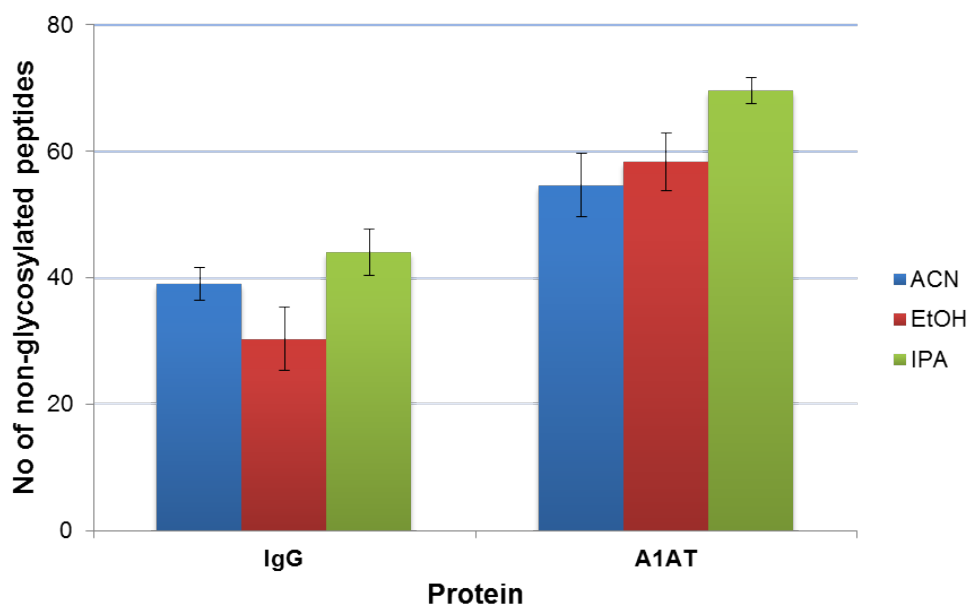


Figure 5.5: Number of identified peptides co-enriched along with the glycopeptides using various solvents. As evident from the results, the number of non-glycosylated peptides co-enriched largely depends upon the mobile phase used.

5.3.6. GLYCOPEPTIDE ENRICHMENT FROM HUMAN SERUM USING DROP-HILIC

Finally, glycopeptide enrichment from human serum before and after depletion of the four most abundant proteins (albumin, A1PI, transferrin and haptoglobin) was evaluated. As observed for the purified standard glycoproteins, the number of enriched glycopeptides varied in a solvent dependant manner (Figure 5.6 A-C). Glycopeptide enrichment efficiencies were determined taking the presence of identified co-enriched, non-glycosylated peptides as an indicator, while a simple automated glycopeptide classification feature available in the ProteinScape software tool was used to estimate the number of enriched glycopeptides. After manual verification of the MS/MS spectra for oxonium ions, just hits with a minimum oxonium ion intensity score of ≥ 60 were considered as glycopeptides. A significant number of relatively low molecular weight, non-glycosylated peptides (between 1000 - 2200 Da) were frequently co-enriched. The degree of non-specific enrichment of higher molecular weight peptides was found to largely dependent on the used solvent. A high number of human serum albumin derived peptides was co-enriched by all loading solvents, but each solvent co-enriched an individual peptide subset (Figure 5.6 D-F). The enriched glycopeptide fractions

were also treated with PNGase F and analysed by RP-nano LC-ESI-MSMS to simply identify enriched, previously glycosylated peptides. The use of acetonitrile, ethanol and isopropanol, respectively, as loading solvent resulted in the identification of 26, 28 and 33 non-redundant, previously glycosylated peptides carrying the deamidated *N*-glycosylation sequence motif. So despite the fact that isopropanol also co-enriched the highest number of unmodified peptides, it also provided the highest number of glycopeptides. These results clearly emphasise that besides (glyco)peptide hydrophilicity sample solvation plays an important role in ZIC-HILIC glycopeptide enrichment.

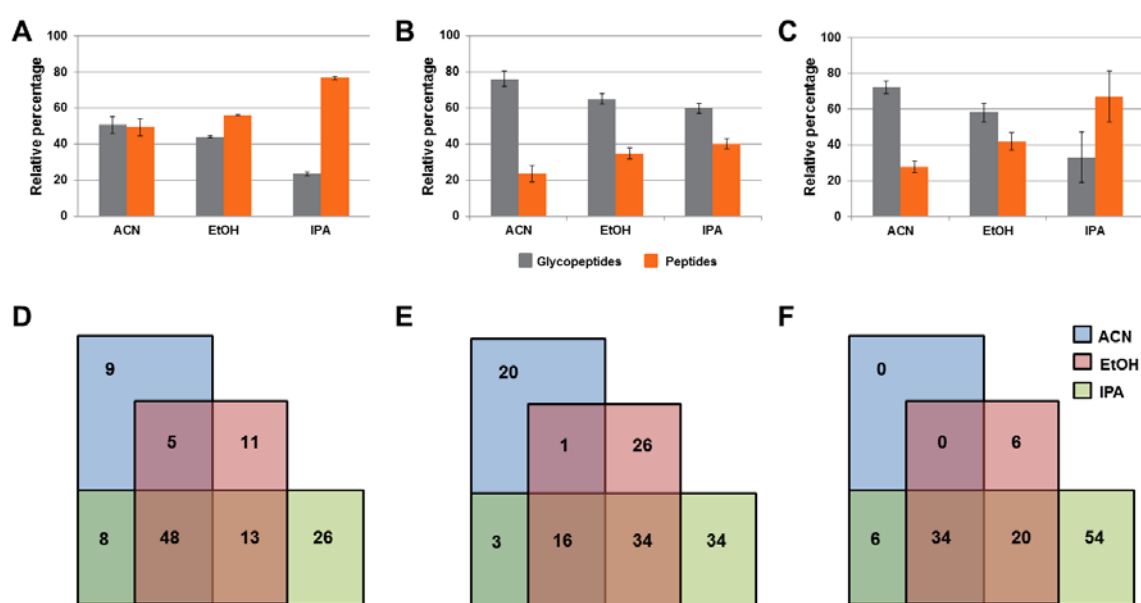


Figure 5.6: Loading solvent effect on ZIC-HILIC glycopeptide enrichment from (A) the four depleted high abundance proteins (B) depleted human serum (C) un-depleted serum. Depending upon the complexity of the sample the relative percentage of the peptides and glycopeptides present in the HILIC enriched fraction varied in a solvent dependent manner. D-F: Venn diagram showing the overlap of various peptides present in the HILIC enriched fraction. Different subset of overlapping and distinct peptides were co-enriched in a solvent dependent manner indicating that, sample solvation plays a major role in HILIC enrichment in addition to hydrophilic partitioning.

5.4. CONCLUSION

With Drop-HILIC a simple, fast and cost effective optimised sample pre-treatment and ZIC-HILIC glycopeptide enrichment strategy was developed. This technique was applied to evaluate the effect the loading solvent has on glycopeptide enrichment efficiency. Independent of whether glycopeptides were enriched from single, purified glycoproteins or complex (glyco)peptide mixtures derived from human serum, ZIC-HILIC glycopeptide enrichment efficiency largely relied on the applied mobile phase but also on the peptide backbone composition. ACN provided the least number of co-enriched peptides while IPA and ethanol showed some preferable features when larger, hydrophobic glycopeptides needed to be enriched. Implementation of orthogonal mobile phase solvents provides one opportunity to increase glycopeptide enrichment efficiency of ZIC-HILIC.

5.5. REFERENCES

1. Kolarich D, Jensen PH, Altmann F, Packer NH. Determination of site-specific glycan heterogeneity on glycoproteins. *Nat Protoc.* 2012;7(7):1285-98.
2. Tang L, Kebarle P. Dependence of Ion Intensity in Electrospray Mass-Spectrometry on the Concentration of the Analytes in the Electrosprayed Solution. *Anal Chem.* 1993;65(24):3654-68.
3. Stavenhagen K, Hinneburg H, Thaysen-Andersen M, Hartmann L, Varon Silva D, Fuchser J, et al. Quantitative mapping of glycoprotein micro-heterogeneity and macro-heterogeneity: an evaluation of mass spectrometry signal strengths using synthetic peptides and glycopeptides. *J Mass Spectrom.* 2013;48(6):627-39.
4. Pasing Y, Sickmann A, Lewandowski U. N-glycoproteomics: mass spectrometry-based glycosylation site annotation. *Biological Chemistry.* 2012;393(4):249-58.
5. Budnik BA, Lee RS, Steen JA. Global methods for protein glycosylation analysis by mass spectrometry. *Biochimica et biophysica acta.* 2006;1764(12):1870-80.
6. An HJ, Froehlich JW, Lebrilla CB. Determination of glycosylation sites and site-specific heterogeneity in glycoproteins. *Current opinion in chemical biology.* 2009;13(4):421-6.
7. Alpert AJ. Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. *J Chromatogr.* 1990;499:177-96.
8. Yoshida T. Peptide separation by Hydrophilic-Interaction Chromatography: a review. *J Biochem Bioph Meth.* 2004;60(3):265-80.
9. Wan H, Yan J, Yu L, Sheng Q, Zhang X, Xue X, et al. Zirconia layer coated mesoporous silica microspheres as HILIC SPE materials for selective glycopeptide enrichment. *The Analyst.* 2011;136(21):4422-30.
10. Snovida SI, Bodnar ED, Viner R, Saba J, Perreault H. 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. *Carbohydrate Research.* 2010;345(6):792-801.
11. Hagglund P, Bunkenborg J, Elortza F, Jensen ON, Roepstorff P. A new strategy for identification of N-glycosylated proteins and unambiguous assignment of their glycosylation sites using HILIC enrichment and partial deglycosylation. *Journal of proteome research.* 2004;3(3):556-66.
12. Takegawa Y, Deguchi K, Ito H, Keira T, Nakagawa H, Nishimura S. Simple separation of isomeric sialylated N-glycopeptides by a zwitterionic type of hydrophilic interaction chromatography. *Journal of separation science.* 2006;29(16):2533-40.
13. Zauner G, Koeleman CA, Deelder AM, Wuhrer M. Protein glycosylation analysis by HILIC-LC-MS of Proteinase K-generated N- and O-glycopeptides. *J Sep Sci.* 2010;33(6-7):903-10.
14. Selman MH, Hemayatkar M, Deelder AM, Wuhrer M. Cotton HILIC SPE microtips for microscale purification and enrichment of glycans and glycopeptides. *Anal Chem.* 2011;83(7):2492-9.
15. Wohlgemuth J, Karas M, Eichhorn T, Hendriks R, Andrecht S. Quantitative site-specific analysis of protein glycosylation by LC-MS using different glycopeptide-enrichment strategies. *Analytical biochemistry.* 2009;395(2):178-88.

16. Wuhrer M, de Boer AR, Deelder AM. Structural glycomics using hydrophilic interaction chromatography (HILIC) with mass spectrometry. *Mass Spectrom Rev.* 2009;28(2):192-206.
17. Ding W, Nothhaft H, Szymanski CM, Kelly J. Identification and Quantification of Glycoproteins Using Ion-Pairing Normal-phase Liquid Chromatography and Mass Spectrometry. *Molecular & Cellular Proteomics.* 2009;8(9):2170-85.
18. Mysling S, Palmisano G, Hojrup P, Thaysen-Andersen M. Utilizing ion-pairing hydrophilic interaction chromatography solid phase extraction for efficient glycopeptide enrichment in glycoproteomics. *Anal Chem.* 2010;82(13):5598-609.
19. Neue K, Mormann M, Peter-Katalinic J, Pohlentz G. Elucidation of glycoprotein structures by unspecific proteolysis and direct nanoESI mass spectrometric analysis of ZIC-HILIC-enriched glycopeptides. *Journal of proteome research.* 2011;10(5):2248-60.
20. Huang J, Guerrero A, Parker E, Strum JS, Smilowitz JT, German JB, et al. Site-specific glycosylation of secretory immunoglobulin A from human colostrum. *Journal of proteome research.* 2015;14(3):1335-49.
21. Wessel D, Flugge UI. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal Biochem.* 1984;138(1):141-3.
22. Piontek C, Varon Silva D, Heinlein C, Pohner C, Mezzato S, Ring P, et al. Semisynthesis of a homogeneous glycoprotein enzyme: ribonuclease C: part 2. *Angew Chem Int Ed Engl.* 2009;48(11):1941-5.
23. Yamamoto N, Ohmori Y, Sakakibara T, Sasaki K, Juneja LR, Kajihara Y. Solid-Phase Synthesis of Sialylglycopeptides through Selective Esterification of the Sialic Acid Residues of an Asn-Linked Complex-Type Sialyloligosaccharide. *Angew Chem Int Ed Engl.* 2003;42(22):2537-40.
24. Kolarich D, Rapp E, Struwe WB, Haslam SM, Zaia J, McBride R, 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):991-5.
25. Taylor CF, Paton NW, Lilley KS, Binz PA, Julian RK, Jr., Jones AR, et al. The minimum information about a proteomics experiment (MIAPE). *Nat Biotechnol.* 2007;25(8):887-93.
26. Hustoft HK, Reubsaet L, Greibrokk T, Lundanes E, Malerod H. Critical assessment of accelerating trypsination methods. *J Pharmaceut Biomed.* 2011;56(5):1069-78.
27. Jiang Z, Smith NW, Liu Z. Preparation and application of hydrophilic monolithic columns. *J Chromatogr A.* 2011;1218(17):2350-61.
28. Gilar M, Olivova P, Daly AE, Gebler JC. Orthogonality of separation in two-dimensional liquid chromatography. *Anal Chem.* 2005;77(19):6426-34.

6. TOWARDS UN-BIASED GLYCOPROTEOMICS - ENHANCING (GLYCO)PEPTIDE IONISATION USING THE CAPTIVESPRAY NANOBOOSTER™

6.1. INTRODUCTION

Nano-LC-ESI-MS/MS has become an indispensable tool in proteomics and glycoproteomics. The significantly reduced flow rate is one key feature of nano LC analyses compared to conventional capillary or analytical flow LC. This reduced flow rate causes emission of smaller droplets from the ESI Taylor cone with higher surface-to-volume ratios allowing enhanced desorption of ions into gas phase that ultimately results in enhanced sensitivity (1). Despite this enhanced sensitivity early nano LC-ESI analyses frequently suffered from an unstable spray resulting in the development of various improved variations of LC-emitter geometries. The CaptiveSpray LC source emitter initially developed by Michrome and then further Bruker Daltonics represents one such development where a vortex gas (e.g. lab air) that sweeps around the LC emitter spray thereby concentrating and focusing the Taylor cone spray into the MS source. The CaptiveSpray emitter also allows for the gas that flows coaxially around the emitter to be enriched with a specific dopant solvent (known as CaptiveSpray nanoBooster™). Depending upon the used dopant, either charge stripping or supercharging of peptides and glycopeptides can be achieved during the ionisation process. Nano-LC-ESI-MS analyses using acetonitrile as dopant showed an increase in peptide ionisation efficiency. During the ESI process hydrophobic molecules tend to be more efficiently ionised compared to hydrophilic ones. Subsequently, glycopeptide signal strengths are significantly lower compared to their unmodified counterparts, mostly due to the presence of the large hydrophilic glycan moiety (2). The CaptiveSpray nanoBooster™ setup using ACN as a dopant, however, was reported to significantly improve glycopeptide ionisation and thus facilitate their detection and analysis (personal communication with Kristina Marx, Bruker). Nevertheless, a systematic study of this phenomenon has to date not been performed.

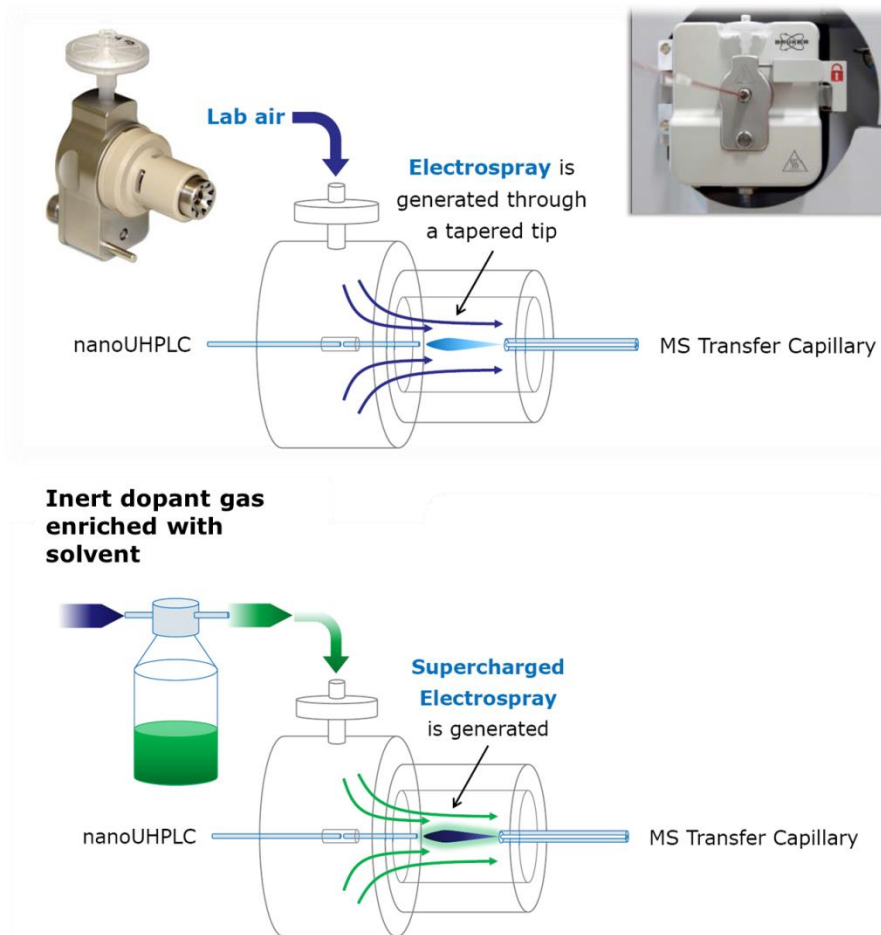


Figure 6.1: Schematic drawing of CaptiveSpray nanoBooster™. (A) In the CaptiveSpray, the vortex gas (e.g. lab air) sweeps around the LC emitter spray thereby concentrating and focusing the Taylor cone spray directly into the MS source. (B) Various dopant enriched nitrogen as sheath gas can be used to enhance the ionisation efficiencies of the hydrophilic analyte compound. In this part of investigation, synthetic glycopeptides was used to systematically investigate the ionisation behaviour of glycopeptides using the CaptiveSpray nanoBooster™ in combination with various MS compatible solvents as dopant (Figure adapted from Bruker Daltonics).

Analyte supercharging using CaptiveSpray nanoBooster™ is particularly advantageous for the ETD fragmentation analysis of glycopeptides due to the reduced mass to charge (m/z) ratio, thereby increasing ETD fragmentation efficiency. However, glycopeptides carrying larger *N*-glycan moieties tend to exhibit low fragmentation yield that is hampering sufficient ETD fragmentation of the peptide moiety and thus, peptide identification. To date the influence of different glycopeptide properties on ETD fragmentation still remains unclear and has not systematically been investigated. Glycopeptide ETD fragmentation provides peptide sequence information while the PTMs still remain intact and attached to the amino acid, also providing crucial site attachment information,

which is in particular of interest for O-glycosylated peptides (3). This particular feature can also be exploited for automated glycopeptide identification using analysis software suits where the PTM can be treated as a variable amino acid modification.

As part of this work synthetic glycopeptides were used to systematically investigate their ionisation behaviour using the CaptiveSpray nanoBooster™ in combination with various MS compatible solvents. In addition, the influence of organic solvent concentrations on the glycopeptide ionisation behaviour was determined besides the various challenges that are associated with glycopeptide ETD fragmentation.

6.2. MATERIALS AND METHODS

If not otherwise stated, all materials were purchased in the highest possible quality from Sigma-Aldrich (St. Louis, MO, USA). Synthetic glycopeptides were synthesised as described previously in the Chapter 4. Nano-LC-ESI-MS analysis was performed as described in the chapter 5 with some minor modification. Briefly, the mass spectrometer was set-up to perform both CID and ETD on the three most intense signals in every MS scan. An m/z range from 400-1600 Da was used for data dependent precursor scanning. The MS data was recorded using the instrument's "enhanced resolution mode". MS/MS data was acquired in "ultra-mode" over an m/z range from 100-3000. Data analysis was performed using ProteinScape 3 (Bruker Daltonics) and MASCOT 2.3 (MatrixScience, United Kingdom) using the following search parameters: Cysteine as carbamidomethyl was set as fixed modification, and deamidation (Asn/Gln) and oxidation (Met) were set as variable modifications. Up to two missed cleavages were allowed. Peptide tolerance was set at ± 0.5 Da for MS and at ± 0.5 Da for MS/MS.

6.3. RESULTS AND DISCUSSION

6.3.1. Rationale and study design

Glycoprotein-focused glycoproteomics aims at acquiring comprehensive data on protein specific glycosylation microheterogeneity (4, 5). For this purpose glycopeptide enrichment is a crucial step for which Hydrophilic Interaction chromatography (HILIC) has extensively been applied due to its low bias towards different glycan types. However, a systematic evaluation of mobile phase effect on

glycopeptide enrichment indicated that ZIC-HILIC glycopeptide efficiency largely relied upon the used solvent (**Chapter 4**). In order to overcome the bias brought in by glycopeptide enrichment, the ionisation behaviour of glycopeptides was investigated immediately after proteolysis and without any prior enrichment by using CaptiveSpray nanoBooster™ ionisation. The influence of various solvents (Acetone, ACN, MeOH, EtOH, and IPA) as well as nitrogen itself to enrich the dopant gas using in glycopeptide ionisation was tested and compared to the results obtained without the use of the nanoBooster™ option.

6.3.2. The quest for the optimal CaptiveSpray nanoBooster™ dopant

The effect various solvents have on glycopeptide ionisation using captive spray nanoBooster™ was first evaluated using a synthetic glycopeptide corresponding to the tryptic peptide sequence derived from IgG2 and carrying bi-antennary *N*-glycan (sample solvent: 30% ACN+0.1% FA) via offline MS. The SPS target mass was set at *m/z* 1350 and any in- or decrease in the signal intensities was expressed in relation to the signal intensity observed when not using the nanoBooster™ option. In addition to glycopeptide signal intensity, parameters such as back ground noise and adduct formation were also considered to determine the optimal solvent for dopant gas enrichment by CaptiveSpray nanoBooster™ ionisation. The presence of multiply charged species of larger analyte molecules is an attractive feature of ESI-MS ionisation, in particular with respect to glycopeptides and their subsequent ETD MS/MS analyses. Initial CaptiveSpray nanoBooster™ experiments using various dopant solvents indicated that the charge distribution (either charge stripping or supercharging) mostly was solvent dependent.

During ESI analyte solution droplets are emitted from the tip of a so called Taylor cone. The emitted droplets undergo rapid solvent evaporation, increasing the charge density up until the point where the cohesive interactions (solvent surface tension) balances the coulombic repulsion (Known as Rayleigh limit). Droplets close to this critical value disintegrate via jet fission and produce highly charged nanodroplets (Figure 6.2). In the case of nanoESI, the small orifice diameter significantly reduces the size of initially produced analyte droplets thereby requiring less evaporation/fission cycles prior MS detection (6-8). The charge state distribution of peptides and proteins was reported to be sample solvent and/or

sheath gas dependent (9, 10). In agreement with previous observations also in the present investigation enhanced glycopeptide ionisation and increased charge states were found to depend on the physio-chemical properties of the solvents such as dielectric constant and surface tension.

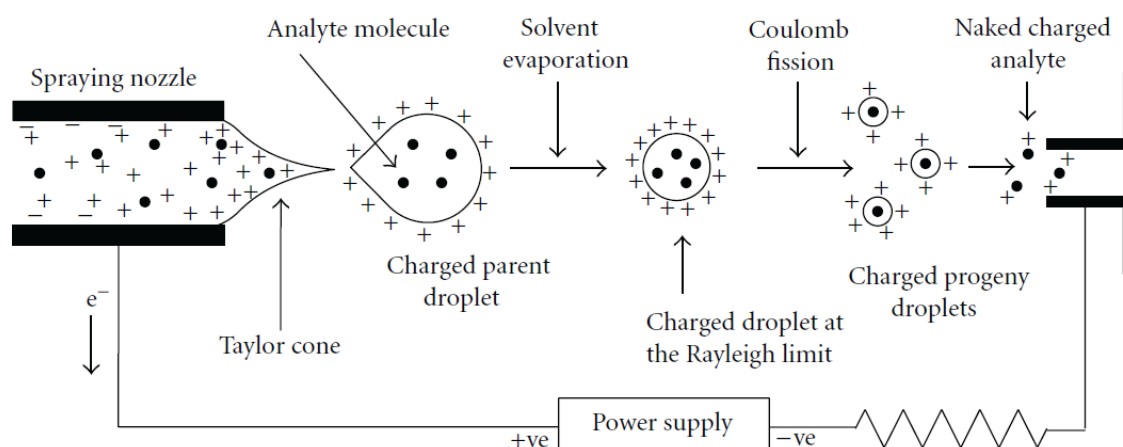


Figure 6.2: Schematic representation of ESI process in positive mode [Figure taken from (11)].

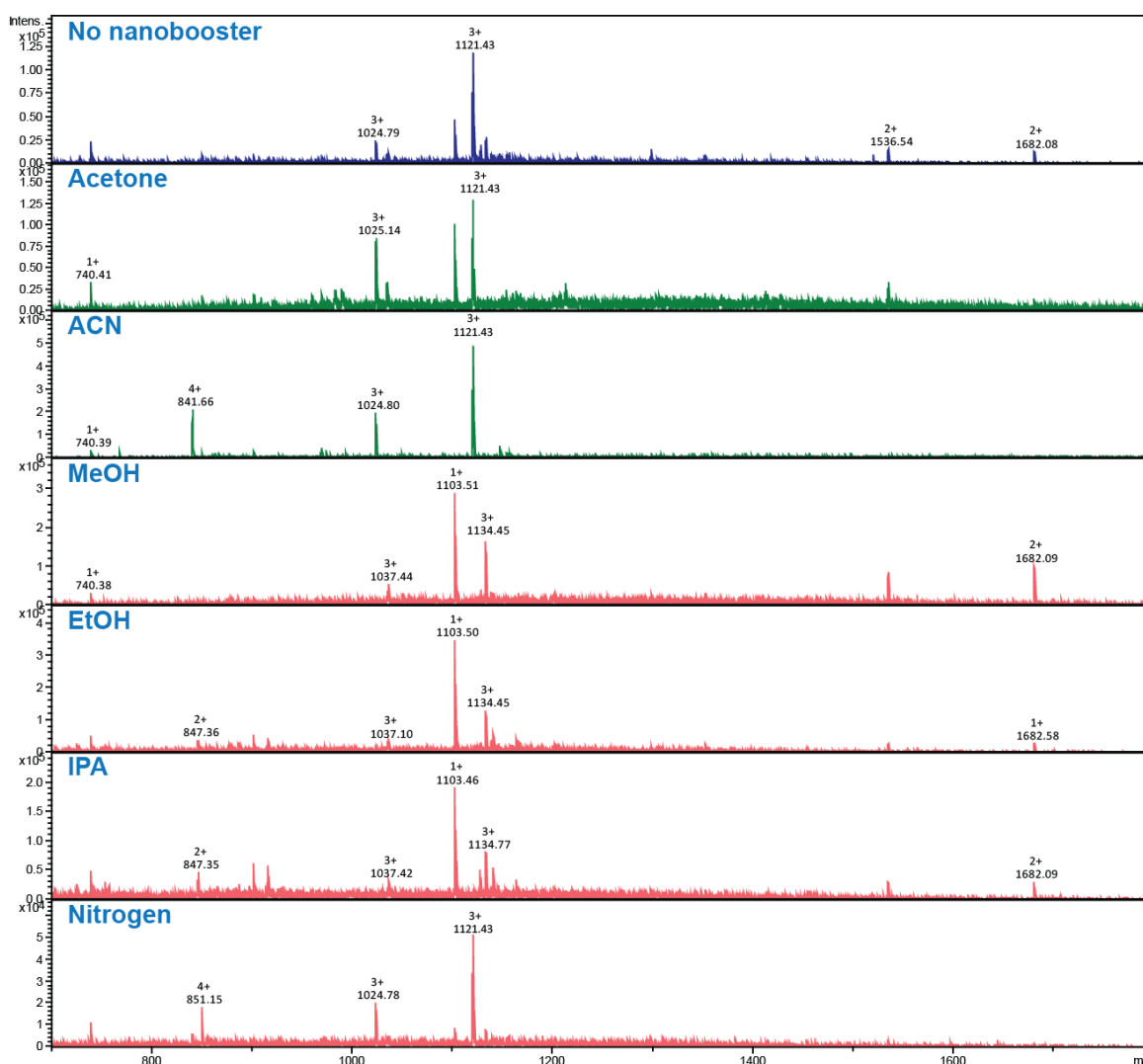
Table 6.1: Physiochemical properties of the used solvents¹

Solvent	Solvent Polarity	Dielectric Constant	Surface tension @ 20°C in mN/m
Acetone	Polar - Aprotic	21	25.20
ACN	Polar - Aprotic	37.5	29.1
MeOH	Polar - Protic	32.6	22.70
EtOH	Polar - Protic	24.3	22.10
IPA	Polar - Protic	18	23.00
Water	Polar - Protic	78.5	72.80
Nitrogen		~1	

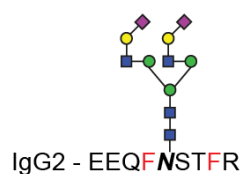
Under the selected MS parameter settings the investigated IgG2 glycopeptides were detected as doubly to quadruply charged molecules, but the intensities of the respective charge states were highly dependent on the used solvents (FIGURE 1). Under conventional ESI-MS analysis conditions, the IgG2 synthetic glycopeptide

¹ Surface tension values were obtained from <http://www.surface-tension.de/> dielectric constants information retrieved from http://chem.libretexts.org/Core/Organic_Chemistry/Fundamentals/Intermolecular_Forces/Polar_Protic_and_Aprotic_Solvents on 02.11.2016

carrying a disialylated, diantennary complex type *N*-glycan was predominately observed as a triply charged signal, and the same result was obtained when nitrogen was used as a vortex gas instead of lab air to focus the Taylor cone spray. Use of MeOH, EtOH and IPA as a dopant solvent resulted in a significant charge stripping as indicated by the increase in the relative signal intensities of the doubly charged precursor ions. Interestingly, these solvents also increased the formation of the potassium adduct for the triply charged ions, whereas no such adduct formation was observed for the doubly charged signals. when acetone and ACN were used as dopant solvents the glycopeptide signal was generally detected in higher charge states. The quadruply charged precursor ions were in particular detected only when ACN was used (Figure 6.3). As ACN exhibited the highest dielectric constant and surface tension compared to the other tested solvents. The observed enhanced glycopeptide ionisation and increased charge states distribution observed when using ACN as dopant confirms the assumption that physio-chemical properties of the solvents determine these features in ESI (Table 6.1, Figure 6.3).



Legend



$$[M + 4H]^{4+} = 841.32$$

$$[M + 3H]^{3+} = 1121.43$$

$$[M + 2H]^{2+} = 1681.64$$

Figure 6.3: Influence of different dopant solvents on the signal intensity and charge state distribution of synthetic IgG2 glycopeptides. Summed up MS spectra over 20 sec elution range are shown with the major peaks labelled accordingly. MeOH, EtOH and IPA as dopant solvents resulted in the enhanced ionisation of low molecular weight contaminants (m/z 1103.51) and the formation of cation adducts (m/z 1134.45) at higher charge states, further complicating glycopeptide analyses. Acetone and ACN as dopant solvents resulted in better signal intensities for the expected analyte as well as lower background noise.

6.3.2.1. Enhanced Glycopeptide ionisation depends upon the solvent used

The influence the tested dopant solvents had on glycopeptide ionisation efficiency was determined by comparing the absolute signal intensities of the detected signals across all charge states. From the tested solvents ACN (5.0X fold), Acetone (3.74-fold), MeOH (2.05-fold) and EtOH (1.24-fold) enhanced glycopeptide signal intensities whereas IPA and nitrogen actually resulted in a decreased signal intensity (0.82-fold and 0.43-fold decrease, respectively) compared to the conventional CaptiveSpray ESI-MS analysis using lab air (Figure 6.4). It is interesting, however, that despite having a higher dielectric constant, MeOH and EtOH has, in comparison to acetone, an opposite effect on glycopeptide ionisation efficiency. This could possibly be explained by the aprotic and protic properties of the solvents. The polar-protic solvents MeOH and EtOH were likely participating in hydrogen bonding with the glycopeptide analyte molecules, thereby reducing their ionisation efficiency. These results clearly supported the assumption that the observed enhanced glycopeptide ionisation was largely depending on the physio-chemical properties of the dopant solvent. Despite the fact that glycopeptide ionisation was just marginally enhanced by MeOH and EtOH, their use as dopant solvents came with additional disadvantages such as increased adduct formation, higher level of background noise especially in the lower mass region and increased ionisation of singly charged contaminant peaks. All these factors together made MeOH and EtOH together with IPA unsuitable dopant solvents for glycoproteomics analyses using the CaptiveSpray nanoBooster™ ionisation device.

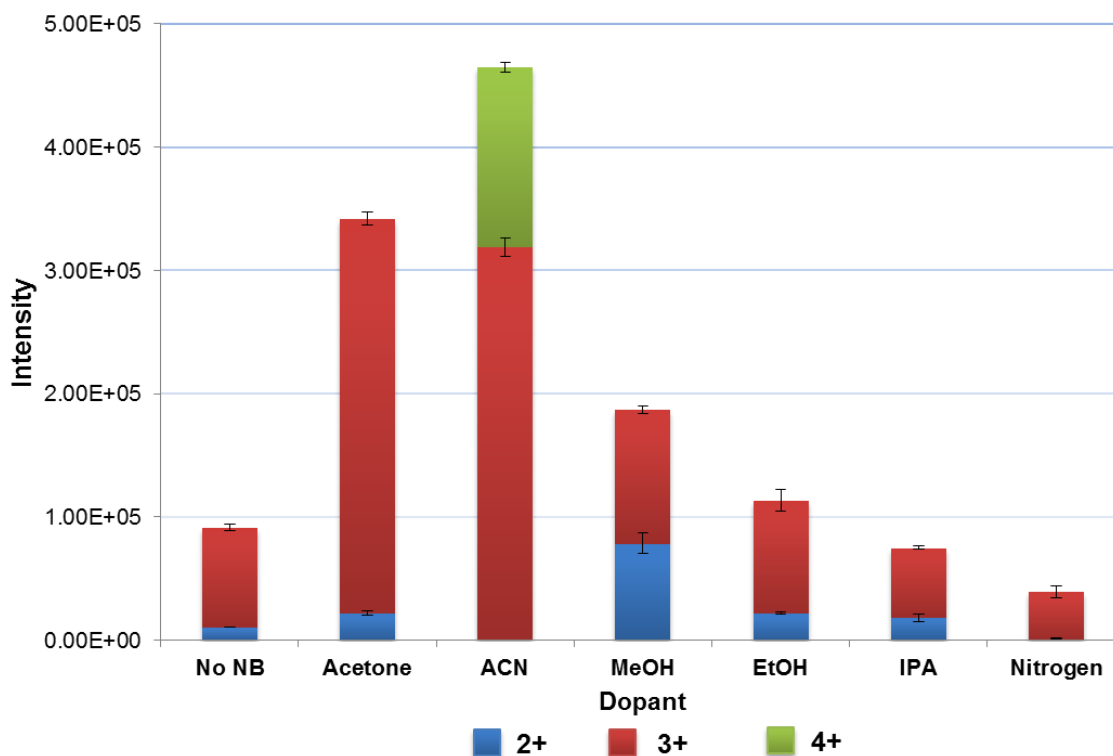


Figure 6.4: Dopant solvent influence on CaptiveSpray nanoBooster™ ionisation as determined on an IgG2 synthetic N-glycopeptide. Among the various tested solvents acetone and ACN were found to be the most suitable dopant solvents. The signal intensities were summed for all detected charge states from MS spectra summed over 20 sec elution range. Mean and standard deviation of triplicate summed up analyses are shown.

The exact molecular mechanisms for the observed charge state enhancement is not known. However, there are several theories that could explain the observed phenomenon. The maximum charge density that an ESI droplet can sustain depends upon the surface tension of the used solvent. During the ionisation process, the size of the droplet decreases, however the number of charges retained by the droplet remains constant, which is resulting in an increased charge density. In simple terms this phenomenon results in reduced charge spacing, thereby increasing the probability that any analyte will be provided with a higher number of charges during the ionization process (Figure 6.5).

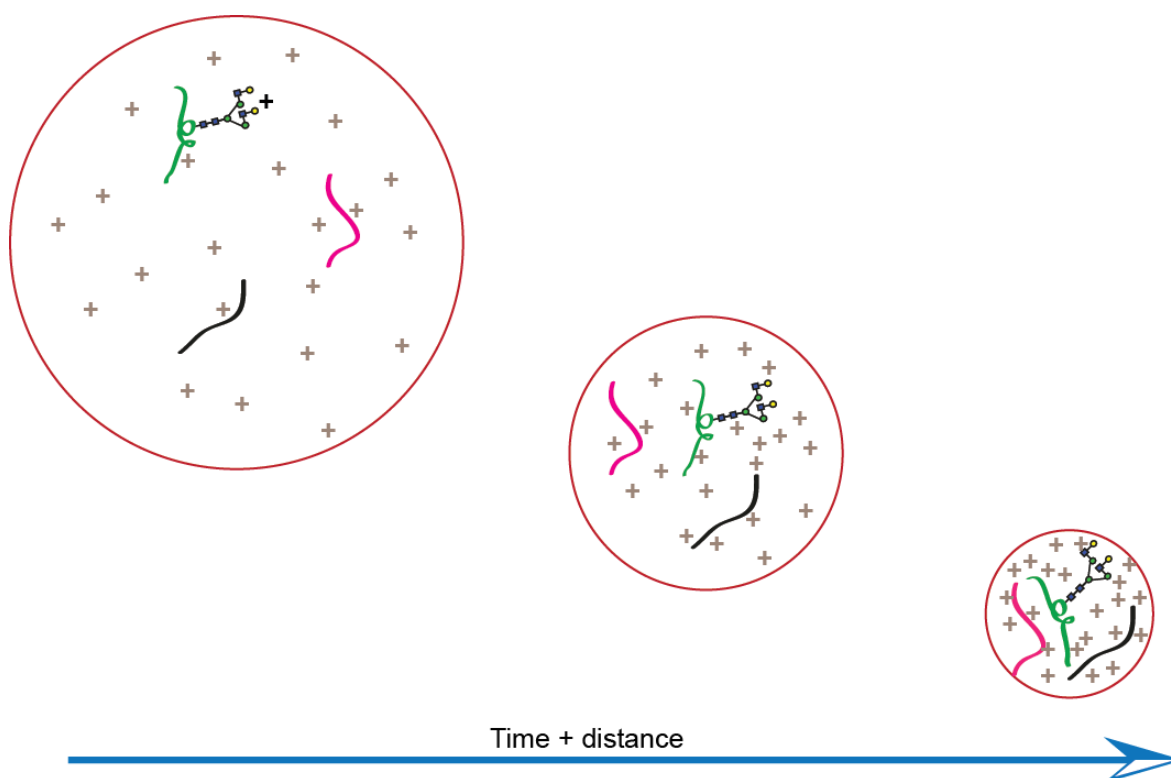


Figure 6.5: During ESI the initial droplet produced at the Taylor cone shrinks until the droplet reaches the Rayleigh limit. The ions evaporated from the resulting nanodroplets are then detected by MS. The use of dopant solvent enriched nitrogen gas during the CaptiveSpray nanoBooster™ ionisation accelerates this process. As a result, a higher charge density is attained much faster than during the conventional n-LC-ESI analysis. As an additional positive side effect this ionisation is also resulting in superior sensitivity.

During CID fragmentation, glycopeptides undergo a preferential fragmentation of the glycan moiety resulting in the appearance of characteristic oxonium ions (B-type ions) in the low m/z region whereas a complete series of Y-type ions corresponding to the consecutive loss of monosaccharides from the non-reducing end dominates the higher m/z region of the MS/MS spectra. The lack or little cleavage of peptide backbone during glycopeptide CID fragmentation is particularly interesting as the peptide backbone can serve as the charge retaining molecule, allowing effective glycopeptide identification via distinct Y1 ions corresponding to the peptide ion carrying a single GlcNAc residue [(peptide+GlcNAc)]. This is due to the fact that the *N*-glycosidic linkage between the core GlcNAc and Asn residue is much stronger compared to the corresponding O-glycosidic bonds present within the glycan. In ion trap instruments, due to intrinsic low-mass cut off filter, trapping of fragment ions obtained during CID fragmentation below 28% of the precursor mass is hampered. This phenomenon is also known as the "1/3 rule". Due to this inherent property of ion trap MS

analysers, diagnostic low-mass oxonium ions (e.g. m/z 204.08; 274.09; 292.10) are often not detected for glycopeptides with a precursor $m/z \geq 1200$. As demonstrated before, the use of the CaptiveSpray nanobooster™ ionisation device enhances glycopeptide ionisation and increases the number of observed charges enabling better detection of these precursors in lower m/z range, and thus also facilitates detection of these low-mass oxonium ions (Figure 6.6). Due to this increased charge state distribution, the observed fragment ions are now also distributed over a narrower m/z range. This property can be further exploited to increase the overall analysis sensitivity by optimising the dynamic MS and MS/MS scan range (Figure 6.7).

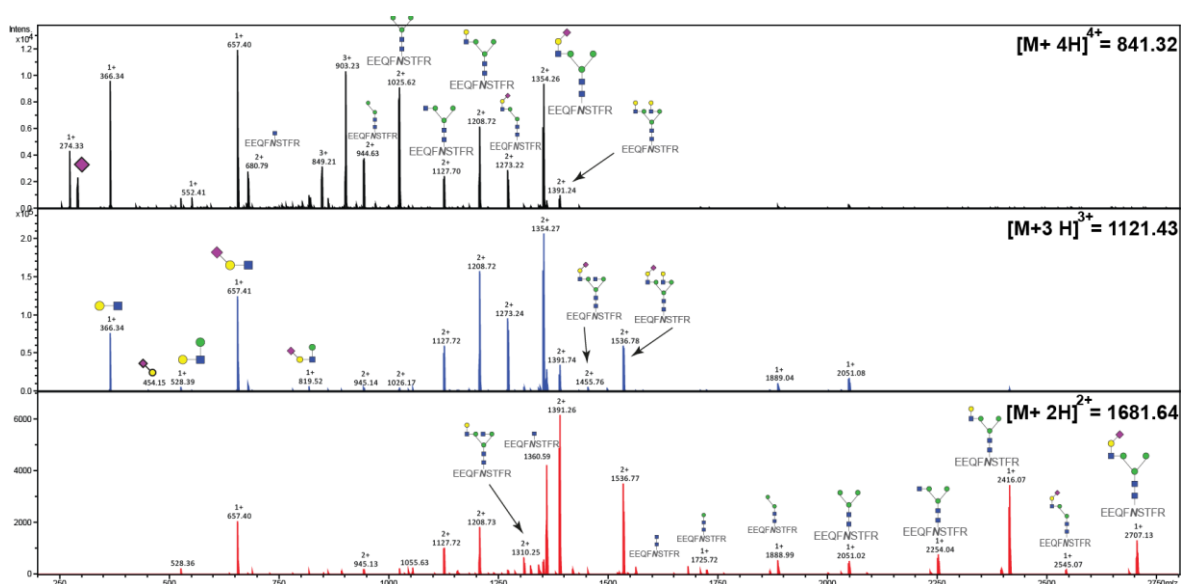


Figure 6.6: CID-MS/MS spectra derived from the synthetic IgG2 glycopeptide carrying a biantennary sialylated N-glycan for (A) quadruply (B) triply and (C) doubly charged precursor ions. Following the $1/3^{\text{rd}}$ rule, the number of observed low-mass oxonium ions decreases significantly with an increase in the precursor m/z .

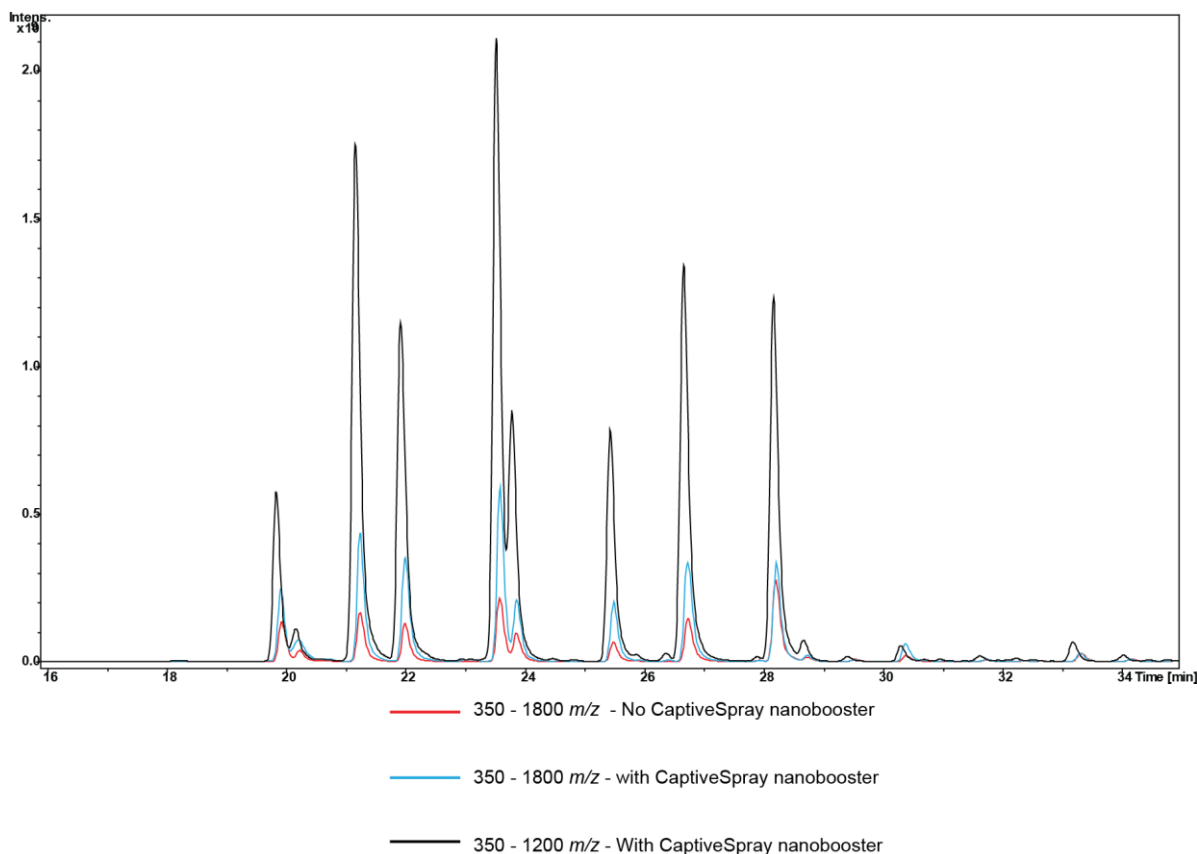


Figure 6.7: Base peak chromatogram of LC-ESI-MS/MS analyses of a Sigma MS retention time calibration mix with and without CaptiveSpray NanoBooster™ ionisation using ACN as the dopant solvent. The use of the CaptiveSpray NanoBooster™ together with an optimised scan range significantly increases the signal intensity of the analyte molecules.

6.3.3. Performance characteristics of ETD glycopeptide fragmentation

In glycoproteomics the combination of two complementary fragmentation techniques such as ETD and CID provides opportunities to characterise the peptide backbone including the site of glycan modification as well as the attached glycan within a single LC-ESI MS/MS experiment (12). Nevertheless, glycopeptide MS/MS data analysis is often a tricky and time-consuming task requiring careful manual interpretation. Depending on the type and size of the modification the resulting spectra can be complex. As mentioned earlier, ETD fragmentation usually results in the preferential cleavage of the peptide backbone while leaving the oligosaccharide portion intact (**Figure 6.8**). The sequence information obtained in ETD fragmentation can potentially be exploited for automated glycopeptide identification using standard proteomics data analysis software where the respective PTMs can be treated as variable modifications. However, the number of variable modifications that reasonably can be included in such a search is limited

to avoid the false positive assignment of too many spectra, and with increasing modification size accurate peak assignment can be also hampered. In order to facilitate software assisted glycopeptide characterisation it is thus necessary to understand the various parameters that are influencing ETD fragmentation. A small synthetic glycopeptide library based on three similar sequences varying in the position of the site of glycosylation was generated and used to systematically investigate parameters that are influencing glycopeptide analysis by ETD fragmentation.

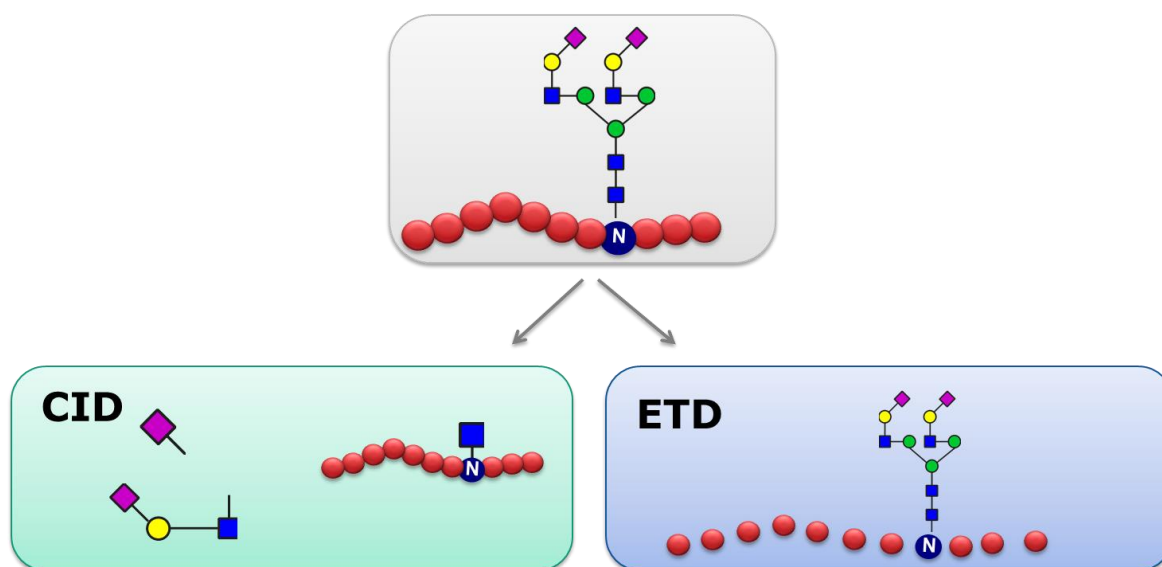
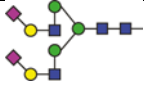
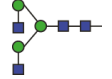
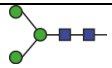


Figure 6.8: Identification of glycopeptides by two complementary fragmentation techniques CID and ETD.

Glycan size and precursor m/z influence ETD glycopeptide fragmentation

The synthetic *N*-glycopeptides exhibiting the peptide sequence ENYSVVFVHPK carrying three different *N*-glycans were subjected to LC-ESI-CID-ETD-MS/MS analysis (Table 6.2).

Table 6.2: Initial set of synthetic glycopeptides used to evaluate the influence of glycan size and precursor m/z on ETD fragmentation

Peptide sequence	N-Glycan	M [Da]	
E <u>N</u> YSVHVHPK		NaN3	3361.2873
		Gn2	2454.9909
		M3	2048.8321

During CID fragmentation, glycopeptides are easily characterised by the presence of diagnostic low-molecular-weight oxonium ions (13). This feature can be used to specifically identify glycopeptides in a large set of MS/MS spectra obtained from an LC-ESI MS/MS analysis. In ETD fragmentation the peptide backbone is fragmented yielding a series of c' ions and z' ion that in part also carry the glycan modification and thus in the best case scenario allow glycosylation site assignment (14). Previously acquired in-house data indicated a precursor m/z and charge state dependence for ETD glycopeptide fragmentation efficiency. For the synthetic glycopeptide ENYSVHVHPK carrying a doubly sialylated, bi-antennary *N*-glycan ETD fragmentation of the triply charged precursor (m/z 1142.12) yielded essentially no information on the peptide sequence (Figure 6.9-A). The quadruple charged precursor (m/z 856.86), however, provided a sufficiently comprehensive z' ions series (z'_{3-8} as singly charged ions and z'_9 as doubly charged ion) allowing peptide identification and assignment of the site of modification (Figure 6.9-B). The complete c -ion series, nevertheless, was totally underrepresented. In the case of this glycopeptide the mass difference of 2318.82 Da between z'_8 and z'_9 indicated the presence of an *N*-glycan corresponding to the composition Hex₅HexNAc₄NeuAc₂.

Next, influence of glycan size on ETD fragmentation was evaluated using the synthetic glycopeptide ENYSVHVHPK carrying Gn2 and M3 *N*-glycan. In both cases, the triply charged precursors (m/z 840.02 and 704.64, respectively) were selected for the subsequent MS/MS analyses as the quadruply charged precursors were not detected in the applied MS scan range. In contrast to the glycopeptide carrying doubly sialylated, biantennary *N*-glycan sufficient data on

the c- and z' ion series could be obtained from the triply charged precursors to confirm the peptide backbone and site of modification (Figure 6.9-C and D).

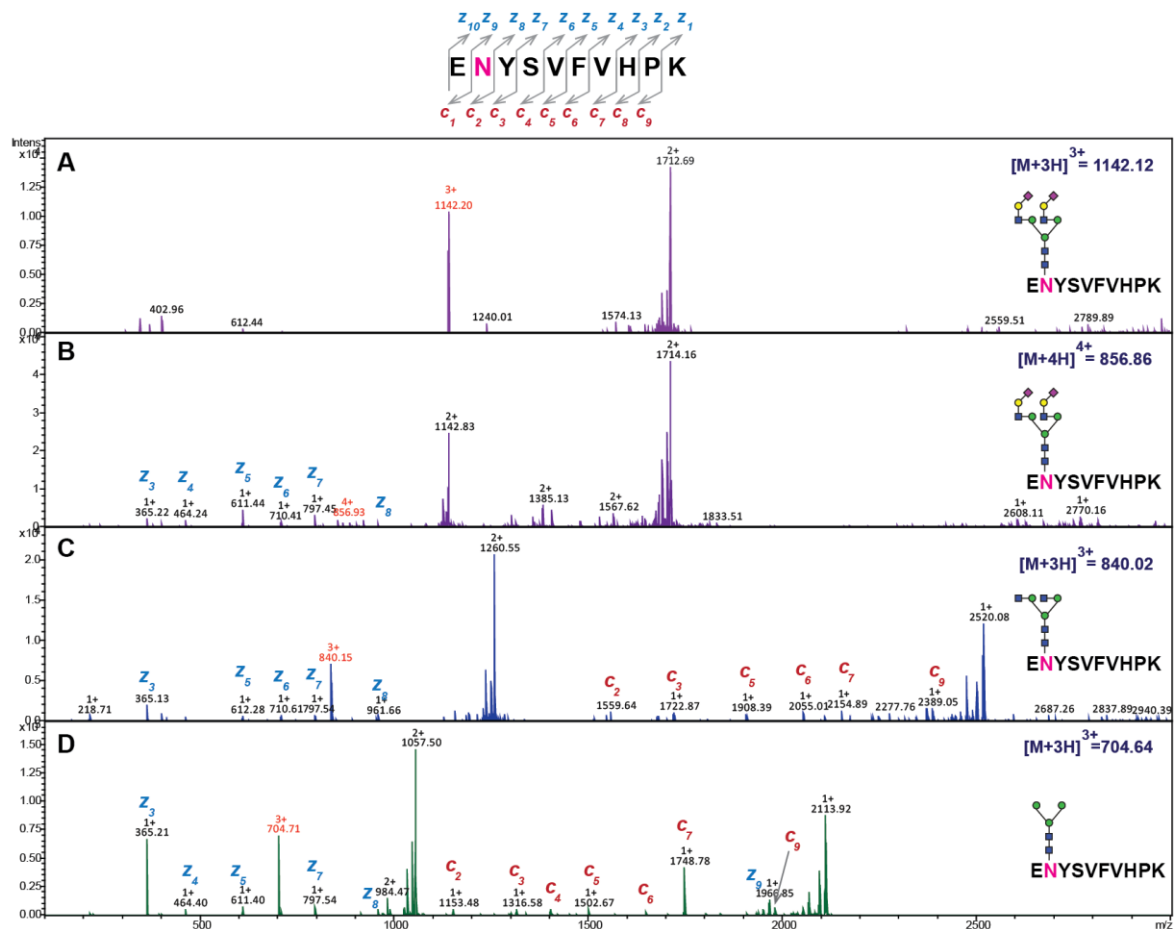


Figure 6.9: ETD MS/MS spectra of a synthetic N-glycopeptide that carries the glycosylation site close to N-terminus (ENYSVFVHPK). Three different glycoforms of this glycopeptide were analysed: (A-B) NaNa, (C) GnGn and (D) M3 N-glycan. A direct correlation was determined between the glycan size and the number of detected c and z ions. In addition to the glycan size the precursor m/z also showed a significant influence on degree of information that could be obtained on the glycopeptide backbone by ETD fragmentation as demonstrated by MS/MS spectra for triply charged precursor ion at m/z 1142.12 (A) and quadruply charged precursor ion at m/z 856.86 (B) as exemplified for the glycopeptide carrying the NaNa N-glycan.

It remains to be seen if and how the sialic acids possibly influence these observed ETD fragmentation patterns in the case of other synthetic glycopeptide described in the chapter 4. In summary the observed results indicated that for glycopeptides ETD fragmentation efficiency (based upon the number of detected c- and z- series ions) largely depended on the precursor m/z and the size of the glycan attached notwithstanding the fact that highly charged precursor ions ($\geq 3+$) were necessary for efficient fragmentation. Under conventional electrospray conditions most N-linked glycopeptides are detected in the m/z range above 900, which is impairing

successful ETD glycopeptide fragmentation for many of these compounds. This drawback, however, can be overcome by the use of improved ionisation devices such as the previously described CaptiveSpray nanoBooster™, that has been demonstrated to enhance glycopeptide ionisation and the number of positive charges on these compounds (see also section XX).

Next, a glycopeptide library based on the peptide sequence ENYSVHVHPK varying only in the site of glycosylation was generated and analysed by LC-ESI-CID-ETD-MS/MS. This glycopeptide library was used to evaluate the influence the site of modification within the glycopeptide sequence has on the ETD fragmentation analysis. The influence the glycosylation site position within a peptide sequence has on ETD fragmentation was evaluated first using three glycopeptides carrying the chitobiose core pentasaccharide (Figure 6.10). In all the three cases, the triply charged precursor ions were selected for ETD fragmentation. The data suggested that the position of the glycosylation site within the sequence influenced the detected length of continuous stretches of c- and z' ions. The presence of a bulk glycan modification near the peptide N-terminus resulted in numerous c and z' ions that could be detected more in the conveniently detectable *m/z* range between 400-1500 Da whereas the presence of the same modification near the peptide's C-terminus or in towards the middle of the sequence pushed these ions further out to the border regions of the ideal MS-scan range, at least on the used instrument. This effect, however, poses a limit on what fragment ions can efficiently be trapped following fragmentation and subsequently detected to provide the diagnostic c and z ion series for peptide sequence determination. Nevertheless, in this particular example the ETD MS/MS spectra still provided useful data to locate the site of modification irrespective of the position of the glycosylation site, but this could also be attributed to the size of the *N*-glycan, as described in detail in the following paragraph.

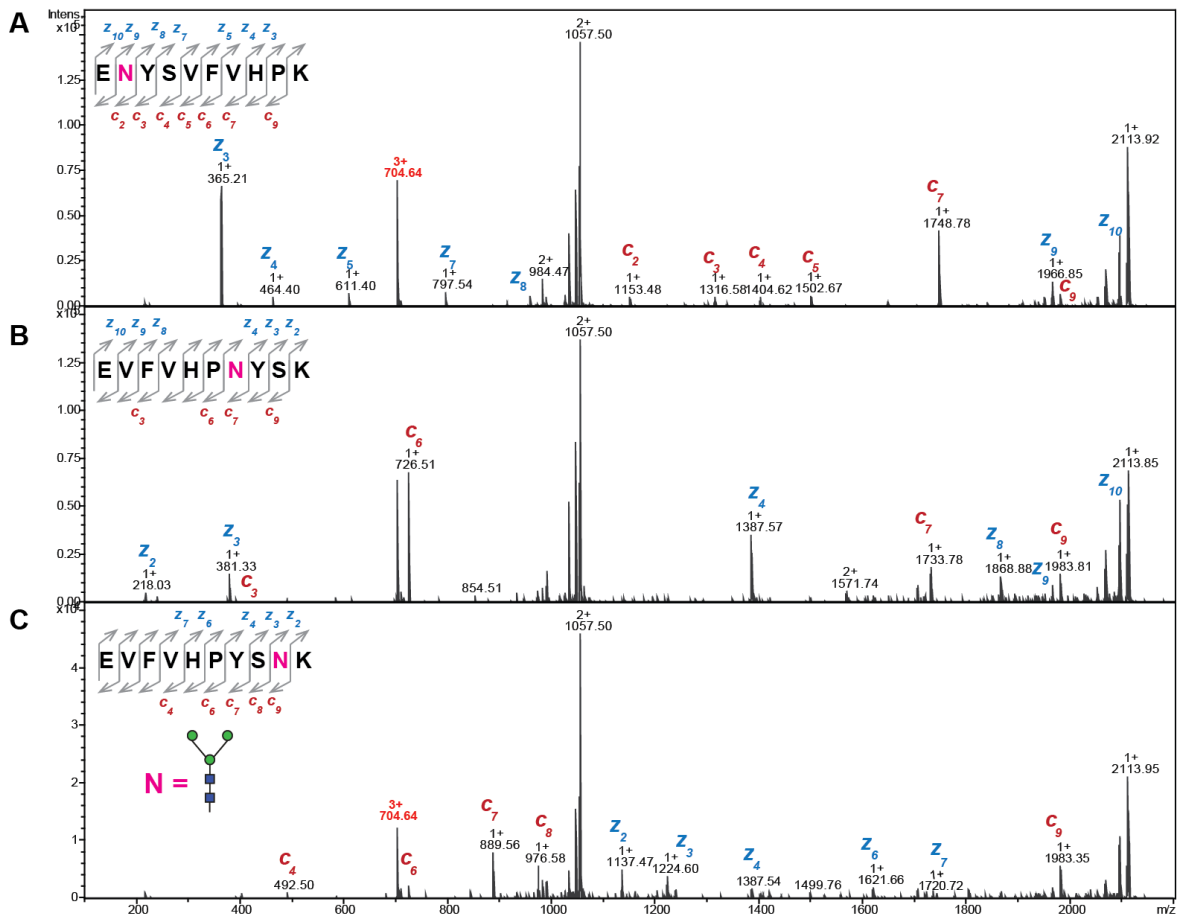


Figure 6.10: Influence of the glycosylation site location within the sequence on ETD fragmentation. ETD MS/MS spectra of three isobaric glycopeptides all carrying the chitobiose core pentasaccharide N-glycan, but differing in the glycosylation site position: (A) N-terminal (B) Middle (C) C-terminal. The position of the glycosylation site within a peptide sequence had a significant effect on the ETD fragmentation efficiency and the number of detected c and z ions, which subsequently are necessary for unambiguous peptide identification.

6.3.3.1. Glycan size and glycosylation site position matter in software assisted glycopeptide identification

Various database search algorithms can be used to search proteomics data, and they have developed to essential proteomic tools required for high throughput automated peptide identification. Traditional software tools used for peptide identification (such as Mascot) match the experimentally observed peptide molecular ion fragments against theoretically generated molecular ions that are derived from a protein sequence database to produce a virtual peptide spectrum and these determined mass values are matched with the theoretically obtained ones (PSM) (15, 16). The more detected m/z signals agree with the theoretical ones, the higher is the probability that a peptide was successfully identified.

Additional factors such as mass accuracy of the detected fragment ions are also included in the various sophisticated peptide identification algorithms and it would go beyond this thesis to elaborate on these too much, but what all these software tools require in the first place is reliable and good experimental data to enable some form of identification.

Most software tools allow to include a relatively small fixed mass offset in the search parameters to accommodate possible sample preparation derived peptide modifications such as carbamidomethylation, oxidation or deamidation without negatively influencing successful peptide identification. To gain a better understanding on the performance characteristics of glycopeptide ETD fragmentation, the Mascot search algorithm, a widely distributed proteomics search tool, was evaluated for its ability to successfully identify *N*-glycopeptides from ETD data. In a first step, the obtained CID-ETD fragmentation spectra obtained for the synthetic glycopeptides were searched against a custom database using Mascot version 2.3, where a set glycan modification was considered as a possible variable modifications. The Mascot Score was used as an indicator for successful glycopeptide identification from the ETD data. As expected, the score depended on the total number and the continuous sequence stretches of detected *c* and *z* ions (see Figure 6.11-A and -B for triply and quadruply charged precursor ions, respectively). Irrespective of the glycosylation site position no successful Mascot identification was achieved from the triply charged precursor ions of the glycopeptides carrying the doubly sialylated, biantennary *N*-glycan (Figure 6.11-A). As no useful diagnostic peptide fragments were obtained from this precursor, this result is not unexpected (Figure 1-A). The glycopeptides carrying the core pentasaccharide or the biantennary, HexNAc terminated *N*-glycans, however, could successfully be identified by the applied Mascot search algorithm, also from the triply charged precursor. The data suggested that the determined Mascot scores were largely depending on the glycan size and location of the glycan modification within the peptide sequence (Figure 6.11). These data clearly show that successful software assisted glycopeptide identification from ETD spectra using Mascot (and possibly also other search algorithms) was highly influenced by precursor *m/z*, glycan size and glycan position within the peptide sequence.

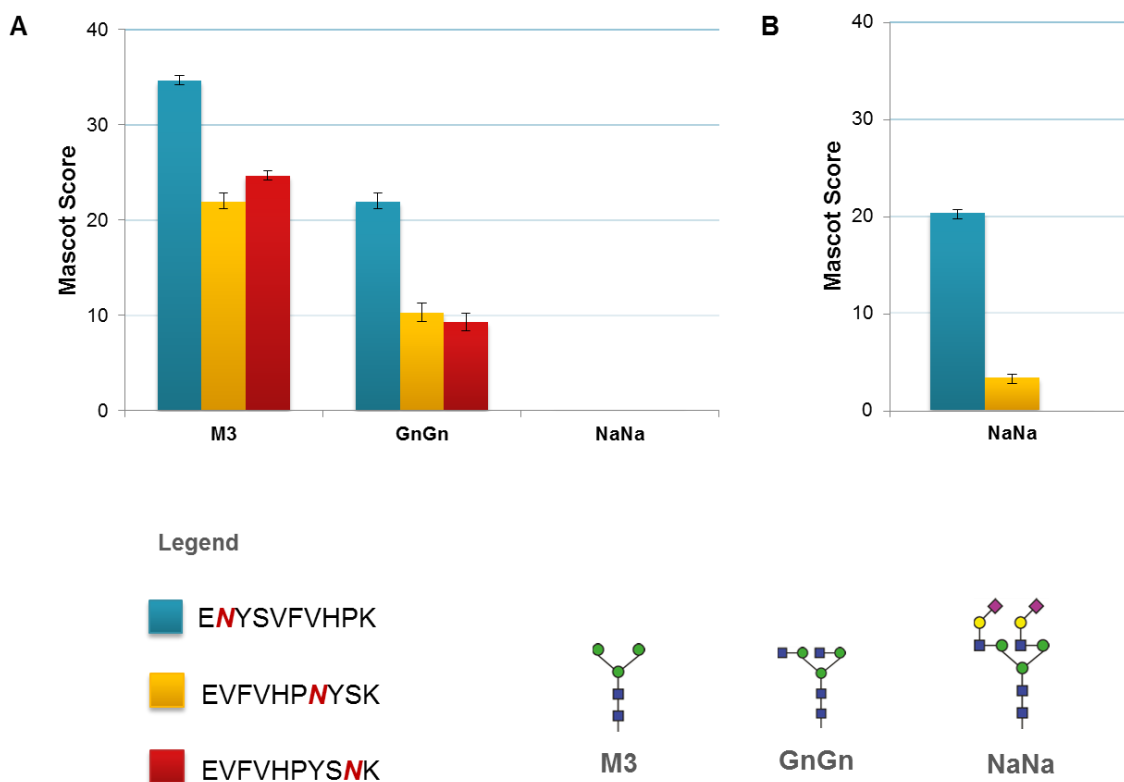


Figure 6.11: Mascot ion scores determined for the synthetic glycopeptide compounds. Glycopeptides differing in N-glycan size were analysed using LC-ESI-CID-ETD MS/MS and searched against a custom made protein database using Mascot 2.3. The search settings used were ± 0.5 Da mass accuracy window for precursor mass and ± 0.5 Da for fragment masses. Values represent the Mascot scores of at least three independent measurements including the \pm standard variation. Mascot scores of triply (A) and quadruply (B) charged precursor ions.

6.4. CONCLUSIONS

The panel of synthetic N-glycopeptides provided an opportunity for a systematic evaluation, glycopeptide analysis method optimisation but fostered also the development of new glycoproteomics approaches. Charge state enhancement and increased enhanced ionisation efficiency of hydrophilic compounds such as glycopeptides could be achieved using a novel ionisation interface, the CaptiveSpray nanoBooster™. The ionisation boosting effect was found to be depending on the used dopant solvent and ACN provided the best properties for enhanced glycopeptide ionisation and supercharging. The synthetic N-glycopeptides were also used to critically evaluate various parameters that are influencing glycopeptide ETD fragmentation with the aim to obtain a better understanding of the processes resulting in glycopeptide fragmentation and to provide a solid basis for data acquisition that resulted in improved software

assisted identification of glycopeptides. The number and quality of peptide backbone fragments detected following glycopeptide ETD fragmentation of glycopeptides were found to depend on glycan size, position of the modification within a peptide sequence and the individual precursor m/z . As a general rule of thumb it was found that highly charged glycopeptides ($z \geq 3$) with precursor masses of $m/z < 900$ significantly facilitated successful ETD fragmentation and hence software assisted glycopeptide identification using the Mascot search algorithm. Future work will show how the instrumental and analytical parameters can be further optimised to push the majority of glycopeptides into that optimal range, so that large scale glycopeptide identification by ETD fragmentation can deliver even more information on the glycoproteome of complex samples.

6.5. REFERENCE

1. Gaspari M, Cuda G. Nano LC-MS/MS: a robust setup for proteomic analysis. *Methods Mol Biol.* 2011;790:115-26.
2. Stavenhagen K, Hinneburg H, Thaysen-Andersen M, Hartmann L, Silva DV, Fuchser J, et al. Quantitative mapping of glycoprotein micro-heterogeneity and macro-heterogeneity: an evaluation of mass spectrometry signal strengths using synthetic peptides and glycopeptides. *J Mass Spectrom.* 2013;48(6):627-39.
3. Klatt S, Rohe M, Alagesan K, Kolarich D, Konthur Z, Hartl D. Production of glycosylated soluble amyloid precursor protein alpha (sAPPalpha) in *Leishmania tarentolae*. *J Proteome Res.* 2013;12(1):396-403.
4. Kolarich D, Jensen PH, Altmann F, Packer NH. Determination of site-specific glycan heterogeneity on glycoproteins. *Nat Protoc.* 2012;7(7):1285-98.
5. Stavenhagen K, Hinneburg H, Thaysen-Andersen M, Hartmann L, Varon Silva D, Fuchser J, et al. Quantitative mapping of glycoprotein micro-heterogeneity and macro-heterogeneity: an evaluation of mass spectrometry signal strengths using synthetic peptides and glycopeptides. *J Mass Spectrom.* 2013;48(6):627-39.
6. Kebarle P, Verkerk UH. Electrospray: from ions in solution to ions in the gas phase, what we know now. *Mass Spectrom Rev.* 2009;28(6):898-917.
7. Wilm MS, Mann M. Electrospray and Taylor-Cone Theory, Does Beam of Macromolecules at Last. *Int J Mass Spectrom.* 1994;136(2-3):167-80.
8. Wilm M, Mann M. Analytical properties of the nanoelectrospray ion source. *Anal Chem.* 1996;68(1):1-8.
9. Kebarle P, Tang L. From Ions in Solution to Ions in the Gas-Phase - the Mechanism of Electrospray Mass-Spectrometry. *Anal Chem.* 1993;65(22):A972-A86.
10. Loo JA, Udseth HR, Smith RD. Solvent Effects on the Charge-Distribution Observed with Electrospray Ionization-Mass Spectrometry of Large Molecules. *Biomed Environ Mass.* 1988;17(5):411-4.
11. Banerjee S, Mazumdar S. Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte. *Int J Anal Chem.* 2012.
12. Alley WR, Jr., Mechref Y, Novotny MV. Characterization of glycopeptides by combining collision-induced dissociation and electron-transfer dissociation mass spectrometry data. *Rapid Commun Mass Spectrom.* 2009;23(1):161-70.
13. Geyer H, Geyer R. Strategies for analysis of glycoprotein glycosylation. *Biochim Biophys Acta.* 2006;1764(12):1853-69.
14. Syrstad EA, Turecek F. Toward a general mechanism of electron capture dissociation. *J Am Soc Mass Spectrom.* 2005;16(2):208-24.
15. Steen H, Mann M. The ABC's (and XYZ's) of peptide sequencing. *Nat Rev Mol Cell Biol.* 2004;5(9):699-711.
16. Eng JK, Searle BC, Clauser KR, Tabb DL. A Face in the Crowd: Recognizing Peptides Through Database Search. *Mol Cell Proteomics.* 2011;10(11).

7. APPENDIX

For Chapter 2: IN-DEPTH STRUCTURAL CHARACTERISATION OF N- GLYCANS USING PGC-NANO-LC-ESI-MS/MS

Appendix Table 1.1: Self-evaluation of the GlycoRRT MS/MS library.

#	RT [min]	Prec. m/z	Pre c. z	MW	Chemical Formula	Compound Name	Cmpd. Comment	Spec. Comment	Lib. RT [min]	RFit'
1 6	31.8	731.2 5	2-	1464 .51	731.27	GnGnF	(HexNAc)2 (Deoxyhexose)1 + (Man)3(GlcNAc)2	D ion 526 and D-18 ion 508 ; m/z 350 for core Fuc; F ion 262 GlcNAc	31.9	1000
5 2	27.6	739.2 5	2-	1480 .51	739.27	GnA	(Hex)1 (HexNAc)2 + (Man)3(GlcNAc)2	D ion 526 and D-18 ion 508	27.7	999
2 2	30.6	820.2 7	2-	1642 .55	820.29	AA	(Hex)2 (HexNAc)2 + (Man)3(GlcNAc)2	D ion 688 and D-18 ion 670	30.7	999
1	27	861.3 2	2-	1724 .66	861.32	(GnGn)(GnGn)	(HexNAc)4 + (Man)3(GlcNAc)2	E ion 507 for (GlcNAc)2; F ion 262 for GlcNAc	27.2	919

For Chapter 5: IT'S ALL ABOUT THE SOLVENT: ON THE IMPORTANCE OF THE MOBILE PHASE FOR ZIC -HILIC GLYCOPEPTIDE ENRICHMENT

Appendix Table 2.1: List of IgG glycopeptides identified after HILIC enrichment.

GP ID	m/z	RT (min)	Z	Calculated GP mass	Glycan	Peptide sequence	GP Abbreviation	Glycan Name
1	1028.89	17.1	3	3083.72	(HexNAc)2 (Deoxyhexose)1 + (Man)3(GlcNAc)2	168-180 TKPREEQFNSTFR	IgG2-GP1	H3N4F1Na0
2	1082.89	17.3	3	3245.84	(Hex)1 (HexNAc)2 (Deoxyhexose)1 + (Man)3(GlcNAc)2	168-180 TKPREEQFNSTFR		H4N4F1Na0
3	1088.21	16.2	3	3261.59	(Hex)2 (HexNAc)2 + (Man)3(GlcNAc)2	168-180 TKPREEQFNSTFR		H5N4F0Na0
4	1096.52	17.3	3	3286.53	(HexNAc)3 (Deoxyhexose)1 + (Man)3(GlcNAc)2	168-180 TKPREEQFNSTFR		H3N5F1Na0
5	1136.84	17.1	3	3407.51	(Hex)2 (HexNAc)2 (Deoxyhexose)1 + (Man)3(GlcNAc)2	168-180 TKPREEQFNSTFR		H5N4F1Na0
6	1150.52	17.4	3	3448.54	(Hex)1 (HexNAc)3 (Deoxyhexose)1 + (Man)3(GlcNAc)2	168-180 TKPREEQFNSTFR		H4N5F1Na0

7	885.15	18.1	4	3536.58	(Hex)1 (HexNAc)2 (Deoxyhexose)1 (NeuAc)1 + (Man)3(GlcNAc)2	168-180 TKPREEQFNSTFR		H4N4F1Na1
8	1234.18	18.2	3	3698.67	(Hex)2 (HexNAc)2 (Deoxyhexose)1 (NeuAc)1 + (Man)3(GlcNAc)2	168-180 TKPREEQFNSTFR		H5N4F1Na1
9	1239.21	16.8	3	3714.6	(Hex)3 (HexNAc)2 (NeuAc)1 + (Man)3(GlcNAc)2	168-180 TKPREEQFNSTFR		H6N4F0Na1
10	1301.74	19.6	2	2601.15	(HexNAc)2 (Deoxyhexose)1 + (Man)3(GlcNAc)2	172-180 EEQFNSTFR	IgG2-GP2	H3N4F1Na0
11	922.14	19.4	3	2764.05	(Hex)1 (HexNAc)2 (Deoxyhexose)1 + (Man)3(GlcNAc)2	172-180 EEQFNSTFR		H4N4F1Na0
12	1403.08	19.6	2	2804.15	(HexNAc)3 (Deoxyhexose)1 + (Man)3(GlcNAc)2	172-180 EEQFNSTFR		H3N5F1Na0
13	1019.09	21.5	3	3054.26	(Hex)1 (HexNAc)2 (Deoxyhexose)1 + (Man)3(GlcNAc)2	172-180 EEQFNSTFR		H4N4F1Na1
14	1073.13	21.5	3	3216.36	(Hex)2 (HexNAc)2 (Deoxyhexose)1 (NeuAc)1 + (Man)3(GlcNAc)2	172-180 EEQFNSTFR		H5N4F1Na1
15	1244.84	16	3	1244.08	(Hex) ₂ (HexNAc) ₂ (Deoxyhexose) ₁ (NeuAc) ₁ + (Man) ₃ (GlcNAc) ₂	172-184 TKPREEQYNSTYR		IgG1-GP3
16	1039.52	15.4	3	3115.54	(HexNAc)2 (Deoxyhexose)1 + (Man)3(GlcNAc)2	172-184 TKPREEQYNSTYR	H3N4F1Na0	
17	1093.54	15.3	3	3278.76	(Hex)1 (HexNAc)2 (Deoxyhexose)1 + (Man)3(GlcNAc)2	172-184 TKPREEQYNSTYR	H4N4F1Na0	
18	1107.19	15.6	3	3318.54	(HexNAc)3 (Deoxyhexose)1 + (Man)3(GlcNAc)2	172-184 TKPREEQYNSTYR	H3N5F1Na0	
19	1147.51	15.2	3	3439.49	(Hex)2 (HexNAc)2 (Deoxyhexose)1 + (Man)3(GlcNAc)2	172-184 TKPREEQYNSTYR	H5N4F1Na0	
20	1161.19	15.4	3	3480.56	(Hex)1 (HexNAc)3 (Deoxyhexose)1 + (Man)3(GlcNAc)2	172-184 TKPREEQYNSTYR	H4N5F1Na0	
21	1190.52	16	3	3568.54	(Hex) ₁ (HexNAc) ₂ (Deoxyhexose) ₁ (NeuAc) ₁ + (Man) ₃ (GlcNAc) ₂	172-184 TKPREEQYNSTYR	H4N4F1Na1	
22	1309.56	18	2	2617.1	(HexNAc)2 (Deoxyhexose)1 + (Man)3(GlcNAc)2	173-181 EEQFNSTYR	IgG4-GP4	H3N4F1Na0
23	1390.57	17.7	2	2779.14	(Hex)1 (HexNAc)2 (Deoxyhexose)1 + (Man)3(GlcNAc)2	173-181 EEQFNSTYR		H4N4F1Na0
24	1317.65	16.4	2	2634.06	(HexNAc)2 (Deoxyhexose)1 + (Man)3(GlcNAc)2	176-184 EEQYNSTYR	IgG1-GP5	H3N4F1Na0
25	932.73	16.3	3	2795.15	(Hex)1 (HexNAc)2 (Deoxyhexose)1 + (Man)3(GlcNAc)2	176-184 EEQYNSTYR		H4N4F1Na0
26	946.41	16.7	3	2836.2	(HexNAc)3 (Deoxyhexose)1 + (Man)3(GlcNAc)2	176-184 EEQYNSTYR		H3N5F1Na0
27	1479.59	16.5	2	2957.16	(Hex)2 (HexNAc)2 (Deoxyhexose)1 + (Man)3(GlcNAc)2	176-184 EEQYNSTYR		H5N4F1Na0
28	1000.4	16.6	3	2998.19	(Hex)1 (HexNAc)3 (Deoxyhexose)1 + (Man)3(GlcNAc)2	176-184 EEQYNSTYR		H4N5F1Na0

Appendix Table 2.2: Comparison of Spin vs Drop HILIC Approach using ACN as loading solvent for IgG. Normalised Data after quantification of the identified glycopeptides

Gp Id		Drop HILIC		Spin HILIC		ACN - Drop			ACN - Spin		
		Mean	SD	Mean	SD	Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3
1	IgG2-GP1	7.07	0.99	4.47	1.63	7.50	5.94	7.78	3.79	6.32	3.29
2		4.47	1.21	2.97	0.68	5.87	3.83	3.72	2.82	3.71	2.38
3		0.48	0.12	0.38	0.11	0.55	0.35	0.55	0.43	0.47	0.26
4		1.09	0.06	0.67	0.12	1.07	1.04	1.15	0.67	0.79	0.56
5		2.99	0.13	2.00	0.56	3.08	2.84	3.05	1.84	2.62	1.55
6		1.54	0.50	0.89	0.57	1.51	1.07	2.06	0.57	1.54	0.55
7		1.70	0.24	1.07	0.14	1.65	1.49	1.96	1.07	1.21	0.92
8		5.03	0.73	3.33	0.67	5.06	4.29	5.74	3.39	3.97	2.62
9		0.60	0.06	0.26	0.05	0.60	0.66	0.54	0.24	0.32	0.23
10	IgG2-GP2	17.86	1.84	19.18	2.44	15.94	19.60	18.02	19.52	16.59	21.43
11		24.67	1.48	28.23	1.22	24.47	26.25	23.31	27.65	27.40	29.63
12		3.02	0.23	3.41	0.20	2.90	3.28	2.87	3.48	3.18	3.57
13		2.29	0.18	2.71	0.09	2.36	2.42	2.08	2.81	2.63	2.68
14		0.99	0.09	1.00	0.11	0.99	0.90	1.08	1.11	0.98	0.90
15	IgG1-GP3	2.24	0.07	2.21	0.09	2.23	2.31	2.18	2.31	2.15	2.16
16		1.56	0.07	1.73	0.07	1.56	1.49	1.63	1.69	1.80	1.69
17		1.36	0.06	1.42	0.14	1.29	1.39	1.40	1.57	1.39	1.30
18		0.38	0.08	0.31	0.07	0.47	0.32	0.35	0.27	0.40	0.28
19		0.81	0.11	0.89	0.06	0.90	0.68	0.85	0.96	0.86	0.86
20		0.64	0.06	0.67	0.04	0.70	0.63	0.58	0.66	0.72	0.63
21		0.19	0.01	0.19	0.02	0.20	0.17	0.19	0.22	0.18	0.19
22	IgG4-GP4	1.21	0.08	1.29	0.10	1.27	1.12	1.25	1.39	1.19	1.30
23		1.41	0.16	1.36	0.06	1.56	1.25	1.43	1.44	1.33	1.33
24	IgG1-GP5	4.40	0.09	5.07	0.23	4.31	4.48	4.43	5.15	4.81	5.24
25		5.66	0.25	7.00	0.48	5.73	5.86	5.38	7.39	6.46	7.15

26		0.81	0.03	0.96	0.07	0.80	0.84	0.79	1.03	0.89	0.95
27		4.36	0.10	4.95	0.20	4.28	4.33	4.47	5.14	4.74	4.97
28		1.16	0.02	1.38	0.02	1.16	1.18	1.15	1.39	1.36	1.40

Appendix Table 2.3: Effect of various incubation times on glycopeptide enrichment efficiency using Drop-HILC approach for IgG. Normalised data after quantification of identified glycopeptides.

Gp Id		Drop 1 min		Drop – 3 min		Drop – 5 min		Drop 1 min			Drop – 3 min			Drop – 5 min		
		Mean	SD	Mean	SD	Mean	SD	Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3
1	IgG2-GP1	7.07	0.99	4.14	2.54	6.25	0.30	7.50	5.94	7.78	1.98	3.51	6.94	6.05	6.60	6.10
2		4.47	1.21	2.64	0.97	3.97	0.39	5.87	3.83	3.72	1.73	2.52	3.66	3.69	4.42	3.80
3		0.48	0.12	0.39	0.08	0.49	0.04	0.55	0.35	0.55	0.30	0.41	0.46	0.53	0.45	0.50
4		1.09	0.06	0.81	0.58	1.15	0.17	1.07	1.04	1.15	0.36	0.62	1.46	1.32	0.98	1.17
5		2.99	0.13	1.91	0.91	2.81	0.23	3.08	2.84	3.05	1.04	1.84	2.85	2.90	2.54	2.98
6		1.54	0.50	0.84	0.69	1.32	0.18	1.51	1.07	2.06	0.30	0.62	1.62	1.44	1.11	1.41
7		1.70	0.24	0.98	0.50	1.49	0.18	1.65	1.49	1.96	0.57	0.84	1.53	1.37	1.42	1.69
8		5.03	0.73	3.63	1.80	4.91	0.10	5.06	4.29	5.74	2.03	3.27	5.58	4.79	4.95	4.98
9		0.60	0.06	0.30	0.08	0.60	0.16	0.60	0.66	0.54	0.21	0.36	0.34	0.75	0.44	0.62
10	IgG2-GP2	17.86	1.84	18.93	2.14	17.25	0.93	15.94	19.60	18.02	21.08	18.90	16.80	18.29	16.50	16.97
11		24.67	1.48	27.69	2.22	25.35	0.49	24.47	26.25	23.31	29.25	28.66	25.15	24.90	25.87	25.27
12		3.02	0.23	3.73	0.56	3.06	0.03	2.90	3.28	2.87	4.24	3.81	3.14	3.10	3.05	3.04
13		2.29	0.18	2.78	0.34	2.35	0.19	2.36	2.42	2.08	3.05	2.90	2.40	2.21	2.56	2.26
14		0.99	0.09	0.91	0.11	0.99	0.01	0.99	0.90	1.08	1.02	0.92	0.80	0.98	1.00	0.99
15	IgG1-GP3	2.24	0.07	2.33	0.20	2.18	0.21	2.23	2.31	2.18	2.10	2.48	2.40	2.14	1.99	2.40
16		1.56	0.07	1.54	0.24	1.62	0.12	1.56	1.49	1.63	1.27	1.62	1.74	1.51	1.60	1.74
17		1.36	0.06	1.27	0.14	1.23	0.15	1.29	1.39	1.40	1.12	1.39	1.31	1.36	1.06	1.26
18		0.38	0.08	0.28	0.04	0.37	0.03	0.47	0.32	0.35	0.25	0.26	0.33	0.34	0.40	0.38
19		0.81	0.11	0.78	0.12	0.71	0.09	0.90	0.68	0.85	0.67	0.90	0.77	0.71	0.63	0.80

20		0.64	0.06	0.64	0.08	0.65	0.04		0.70	0.63	0.58	0.56	0.70	0.68	0.62	0.69	0.63
21		0.19	0.01	0.16	0.03	0.20	0.03		0.20	0.17	0.19	0.12	0.17	0.18	0.22	0.17	0.20
22	IgG4-GP4	1.21	0.08	1.47	0.17	1.32	0.11		1.27	1.12	1.25	1.64	1.45	1.31	1.24	1.45	1.28
23		1.41	0.16	1.57	0.26	1.50	0.15		1.56	1.25	1.43	1.87	1.42	1.42	1.54	1.63	1.34
24	IgG1-GP5	4.40	0.09	5.68	0.83	4.83	0.34		4.31	4.48	4.43	6.50	5.70	4.85	4.49	5.18	4.82
25		5.66	0.25	6.92	1.24	6.24	0.23		5.73	5.86	5.38	8.12	7.00	5.64	6.44	5.99	6.29
26		0.81	0.03	1.02	0.12	0.98	0.15		0.80	0.84	0.79	1.13	1.05	0.89	1.01	1.12	0.82
27		4.36	0.10	5.10	0.62	4.80	0.13		4.28	4.33	4.47	5.72	5.08	4.48	4.72	4.73	4.95
28		1.16	0.02	1.55	0.25	1.37	0.10		1.16	1.18	1.15	1.78	1.59	1.29	1.33	1.48	1.29

Appendix Table 2.4: Effect of various solvents on glycopeptide enrichment efficiency using Drop-HILC approach. Normalised data after quantification of identified glycopeptides.

Gp Id		ACN		EtOH		IPA			ACN			EtOH			IPA		
		Mean	SD	Mean	SD	Mean	SD		Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3
1	IgG2-GP1	7.07	0.99	3.12	0.56	6.56	0.15		7.50	5.94	7.78	3.15	2.54	3.66	6.52	6.43	6.72
2		4.47	1.21	2.46	0.43	4.20	0.16		5.87	3.83	3.72	2.61	1.98	2.80	4.39	4.09	4.12
3		0.48	0.12	0.31	0.04	0.43	0.06		0.55	0.35	0.55	0.35	0.32	0.27	0.44	0.36	0.49
4		1.09	0.06	0.55	0.07	1.45	0.12		1.07	1.04	1.15	0.47	0.60	0.58	1.55	1.32	1.49
5		2.99	0.13	1.78	0.21	2.79	0.29		3.08	2.84	3.05	1.67	1.65	2.03	3.10	2.54	2.74
6		1.54	0.50	0.66	0.17	1.29	0.18		1.51	1.07	2.06	0.52	0.60	0.85	1.12	1.27	1.48
7		1.70	0.24	0.81	0.05	1.61	0.07		1.65	1.49	1.96	0.75	0.83	0.85	1.69	1.55	1.59
8		5.03	0.73	2.81	0.06	4.43	0.40		5.06	4.29	5.74	2.87	2.75	2.80	4.29	4.13	4.88
9		0.60	0.06	0.45	0.05	0.60	0.03		0.60	0.66	0.54	0.48	0.39	0.47	0.57	0.58	0.63
10	IgG2-GP2	17.86	1.84	19.55	0.09	18.96	0.12		15.94	19.60	18.02	19.58	19.62	19.44	18.83	19.07	18.97
11		24.67	1.48	27.56	1.32	23.62	1.19		24.47	26.25	23.31	26.98	29.08	26.62	22.48	24.85	23.52
12		3.02	0.23	3.63	0.06	2.83	0.12		2.90	3.28	2.87	3.66	3.66	3.56	2.96	2.74	2.79
13		2.29	0.18	2.65	0.04	2.24	0.07		2.36	2.42	2.08	2.70	2.64	2.62	2.32	2.23	2.18
14		0.99	0.09	1.17	0.26	1.73	0.07		0.99	0.90	1.08	0.98	1.07	1.47	1.80	1.73	1.67

15	IgG1-GP3	2.24	0.07	2.59	0.26	2.09	0.11		2.23	2.31	2.18	2.29	2.79	2.68	2.20	1.99	2.09
16		1.56	0.07	1.62	0.07	1.64	0.18		1.56	1.49	1.63	1.58	1.58	1.70	1.66	1.45	1.81
17		1.36	0.06	1.60	0.11	1.85	0.38		1.29	1.39	1.40	1.57	1.72	1.51	1.45	1.89	2.20
18		0.38	0.08	0.34	0.03	0.47	0.04		0.47	0.32	0.35	0.31	0.36	0.36	0.51	0.45	0.44
19		0.81	0.11	1.12	0.17	1.00	0.12		0.90	0.68	0.85	1.25	1.16	0.93	1.02	1.10	0.87
20		0.64	0.06	0.59	0.02	0.68	0.14		0.70	0.63	0.58	0.59	0.56	0.61	0.67	0.54	0.82
21		0.19	0.01	0.18	0.02	0.22	0.01		0.20	0.17	0.19	0.16	0.20	0.18	0.20	0.23	0.23
22	IgG4-GP4	1.21	0.08	1.19	0.15	1.02	0.08		1.27	1.12	1.25	1.37	1.10	1.11	0.93	1.04	1.09
23		1.41	0.16	1.52	0.14	1.33	0.13		1.56	1.25	1.43	1.67	1.43	1.45	1.18	1.43	1.37
24	IgG1-GP5	4.40	0.09	5.45	0.21	4.25	0.20		4.31	4.48	4.43	5.52	5.22	5.63	4.48	4.18	4.09
25		5.66	0.25	8.01	0.33	6.40	0.38		5.73	5.86	5.38	8.12	8.26	7.63	6.76	6.45	6.00
26		0.81	0.03	1.07	0.11	0.83	0.16		0.80	0.84	0.79	1.08	1.17	0.96	1.01	0.77	0.71
27		4.36	0.10	5.61	0.33	4.37	0.35		4.28	4.33	4.47	5.96	5.29	5.59	4.61	4.52	3.97
28		1.16	0.02	1.61	0.17	1.13	0.11		1.16	1.18	1.15	1.76	1.43	1.64	1.25	1.09	1.04

Appendix Table 2. 5: List of A1PI glycopeptides identified after HILIC enrichment.

GP ID	m/z	RT (min)	Z	Calculated GP mass	Glycan	Peptide sequence	GP Abbreviation	Glycan Name
1	1388.5 5	40.8	4	5547.36	(Hex)3 (HexNAc)2 (NeuAc)2 + (Man)3(GlcNAc)2	64-93 QLAHQSNSTNIFFSPVSIAT AFAMLSLGTK	A1PI-GP1	H6N4F0Na2
2	1352.0 8	40.9	4	5401.4	(Hex)2 (HexNAc)2 (NeuAc)2 + (Man)3(GlcNAc)2	64-93 QLAHQSNSTNIFFSPVSIAT AFAMLSLGTK		H5N4F0Na1
3	1279.3 8	38.4	4	5110.52	(Hex)2 (HexNAc)2 (NeuAc)1 + (Man)3(GlcNAc)2	64-93 QLAHQSNSTNIFFSPVSIAT AFAMLSLGTK		H5N4F0Na1
4	1275.3 6	39.7	4	5094.56	(Hex)2 (HexNAc)2 (NeuAc)1 + (Man)3(GlcNAc)2	64-93 QLAHQSNSTNIFFSPVSIAT AFAMLSLGTK		H5N4F0Na2
5	1475.6 3	38.1	3	4422.39	(HexNAc)1 (Deoxyhexose)1 + (Man)3(GlcNAc)2	64-93 QLAHQSNSTNIFFSPVSIAT AFAMLSLGTK		H3N3F1Na1
6	1402.8 5	36	4	5604.36	(Hex)2 (HexNAc)2 (NeuAc)1 + (Man)3(GlcNAc)2	94-125 ADTHDEILEGLNFNLTPEIQIHEGFQELLR	A1PI-GP2	H5N4F0Na1
7	1180.8 3	38	5	5896.05	(Hex)2 (HexNAc)2 (NeuAc)2 + (Man)3(GlcNAc)2	94-125 ADTHDEILEGLNFNLTPEIQIHEGFQELLR		H5N4F0Na2
8	1224.6	33	3	3668.7	(Hex)2 (HexNAc)2 (NeuAc)1 + (Man)3(GlcNAc)2	268-283 YLGNATAIFFFLPDEGK	A1PI-GP3	H5N4F0Na1
9	1321.2 9	34.7	3	3959.67	(Hex)2 (HexNAc)2 (NeuAc)2 + (Man)3(GlcNAc)2	268-283 YLGNATAIFFFLPDEGK		H5N4F0Na2
10	1092.1 7	34.6	5	5453.8	(Hex)2 (HexNAc)2 (NeuAc)1 + (Man)3(GlcNAc)2	268-298 YLGNATAIFFFLPDEGKQLQHL ENELTHDIITK (1 missed cleavage)	A1PI-GP4	H5N4F0Na1
11	1150.3 6	35.9	5	5744.95	(Hex)2 (HexNAc)2 (NeuAc)2 + (Man)3(GlcNAc)2	268-298 YLGNATAIFFFLPDEGKQLQHL ENELTHDIITK (1 missed cleavage)		H5N4F0Na2
12	1438.8	36	4	5744.24	(Hex)2 (HexNAc)2 (NeuAc)2 + (Man)3(GlcNAc)2	268-298 YLGNATAIFFFLPDEGKQLQHL ENELTHDIITK (1 missed cleavage)		H5N4F0Na2

Appendix Table 2.6: Comparison of Spin vs Drop HILIC Approach using ACN as loading solvent for A1PI sample. Normalised Data after quantification of the identified glycopeptides

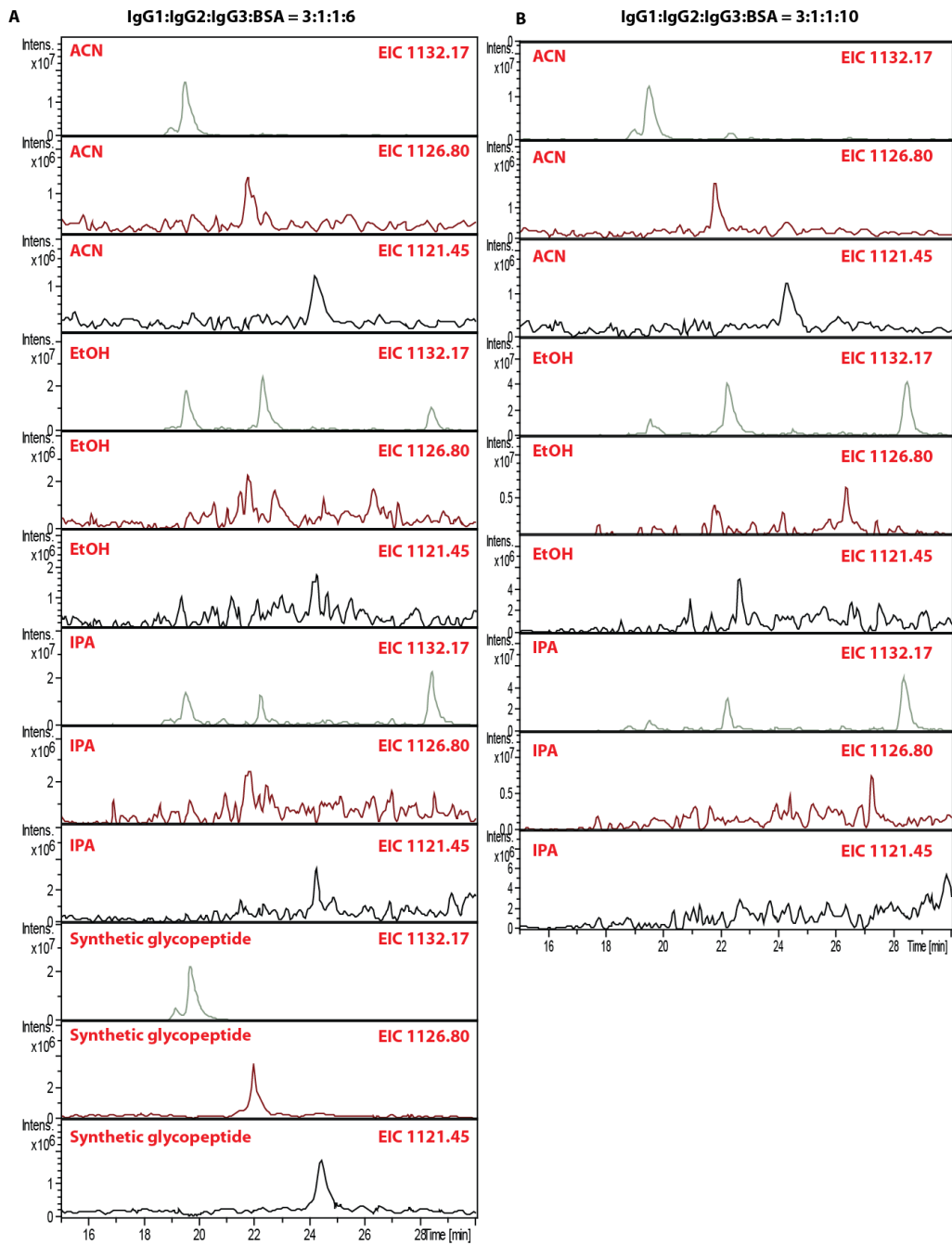
Gp Id		Drop HILIC		Spin HILIC		ACN - Drop			ACN - Spin		
		Mean	SD	Mean	SD	Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3
1	A1PI-GP1	0.00	0.00	0.42	0.12	0.00	0.00	0.00	0.55	0.30	0.41
2		0.38	0.17	1.31	0.21	0.25	0.58	0.32	1.11	1.30	1.52
3		0.00	0.00	0.38	0.10	0.00	0.00	0.00	0.27	0.42	0.46
4		2.16	0.36	9.44	2.34	1.84	2.10	2.54	12.14	8.09	8.08
5		0.71	0.29	4.84	0.64	0.38	0.81	0.94	5.56	4.63	4.32
6	A1PI-GP2	0.31	0.14	0.74	0.29	0.29	0.45	0.18	0.49	0.66	1.06
7		0.30	0.17	1.57	0.19	0.10	0.43	0.37	1.74	1.37	1.60
8	A1PI-GP3	14.84	1.01	7.69	2.29	14.63	15.94	13.96	5.14	8.34	9.58
9		79.04	1.36	62.52	3.20	80.39	77.67	79.05	66.21	60.55	60.80
10	A1PI-GP4	0.00	0.00	0.89	0.52	0.00	0.00	0.00	0.30	1.27	1.10
11		2.26	0.33	10.20	3.38	2.12	2.03	2.64	6.47	13.07	11.06

Appendix Table 2.7: Effect of various incubation times on glycopeptide enrichment efficiency using Drop-HILC approach for A1PI. Normalised data after quantification of identified glycopeptides.

Gp Id		Drop 1 min		Drop – 3 min		Drop – 5 min			Drop 1 min			Drop – 3 min			Drop – 5 min		
		Mean	SD	Mean	SD	Mean	SD		Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3
1	A1PI-GP1	0.00	0.00	0.00	0.00	0.00	0.00		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2		0.38	0.17	0.45	0.12	0.57	0.08		0.25	0.58	0.32	0.39	0.38	0.59	0.66	0.53	0.51
3		0.00	0.00	0.00	0.00	0.00	0.00		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4		2.16	0.36	4.17	1.85	4.88	1.06		1.84	2.10	2.54	3.99	2.41	6.10	5.99	4.77	3.89
5		0.71	0.29	0.54	0.36	2.11	0.92		0.38	0.81	0.94	0.49	0.21	0.92	3.15	1.73	1.43
6	A1PI-GP2	0.31	0.14	0.00	0.00	0.35	0.13		0.29	0.45	0.18	0.00	0.00	0.00	0.50	0.29	0.25
7		0.30	0.17	0.00	0.00	0.76	0.26		0.10	0.43	0.37	0.00	0.00	0.00	1.04	0.73	0.52
8	A1PI-GP3	14.84	1.01	11.14	2.66	9.56	1.02		14.63	15.94	13.96	13.56	8.29	11.57	8.70	9.30	10.69
9		79.04	1.36	83.02	4.72	78.92	1.72		80.39	77.67	79.05	80.85	88.44	79.78	77.06	80.45	79.26
10	A1PI-GP4	0.00	0.00	0.00	0.00	0.00	0.00		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
11		2.26	0.33	0.68	0.39	2.85	0.63		2.12	2.03	2.64	0.72	0.27	1.04	2.89	2.20	3.45

Appendix Table 2.8: Effect of various solvents on glycopeptide enrichment efficiency using Drop-HILC approach for A1PI. Normalised data after quantification of identified glycopeptides.

Gp Id		ACN		EtOH		IPA			ACN			EtOH			IPA		
		Mean	SD	Mean	SD	Mean	SD		Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3
1	A1PI-GP1	0.00	0.00	0.00	0.00	0.73	0.16		0.00	0.00	0.00	0.00	0.00	0.00	0.92	0.63	0.65
2		0.38	0.17	0.49	0.03	1.65	0.17		0.25	0.58	0.32	0.46	0.53	0.49	1.82	1.48	1.64
3		0.00	0.00	0.21	0.02	0.23	0.03		0.00	0.00	0.00	0.19	0.23	0.22	0.21	0.22	0.26
4		2.16	0.36	4.13	0.73	17.35	1.55		1.84	2.10	2.54	3.50	4.93	3.95	19.09	16.84	16.12
5		0.71	0.29	2.45	0.22	9.12	2.03		0.38	0.81	0.94	2.24	2.68	2.44	11.46	8.00	7.89
6	A1PI-GP2	0.31	0.14	0.40	0.05	0.90	0.18		0.29	0.45	0.18	0.38	0.46	0.35	0.99	0.69	1.01
7		0.30	0.17	0.77	0.01	2.30	0.50		0.10	0.43	0.37	0.79	0.77	0.77	2.86	2.12	1.91
8	A1PI-GP3	14.84	1.01	7.69	0.61	5.17	0.55		14.63	15.94	13.96	7.09	7.68	8.30	4.64	5.15	5.73
9		79.04	1.36	80.74	1.70	57.63	4.49		80.39	77.67	79.05	82.42	79.01	80.80	52.79	61.65	58.46
10	A1PI-GP4	0.00	0.00	0.00	0.00	0.24	0.03		0.00	0.00	0.00	0.00	0.00	0.00	0.24	0.20	0.27
11		2.26	0.33	3.10	0.54	4.69	1.54		2.12	2.03	2.64	2.94	3.70	2.67	5.00	3.03	6.06



Appendix Figure 2.1 (A-B): Glycopeptide enrichment efficiency evaluation using the synthetic glycopeptide. When Excess molar ratio of BSA was applied, the enrichment efficiency was significantly compromised in the case of EtOH and IPA as evident from the EIC of the synthetic glycopeptides

8. Curriculum Vitae

For reasons of data protection, the Curriculum vitae is not published in the online version.

