Targeting the human glycoproteome

New enrichment protocols and mass spectrometric analyses reveal unique and novel glycosylation sites

Adnan Halim

Department of Clinical Chemistry and Transfusion Medicine
Institute of Biomedicine
Sahlgrenska Academy at the University of Gothenburg

UNIVERSITY OF GOTHENBURG
Gothenburg 2012

Targeting the human glycoproteome

adnan.halim@clinchem.gu.se

ISBN 978-91-628-8412-3
Abstract and thesis frame available online:
http://hdl.handle.net/2077/28244

Printed by Ineko AB, Gothenburg, Sweden 2012

ABSTRACT

Glycosylation is one of the most common and structurally diverse post-translational modifications of proteins. Given that protein glycosylation is involved in various cellular processes, the characterization of site-specific N- and O-linked glycosylations is becoming increasingly important. However, current mass spectrometry-based technologies, i.e. proteomics and glycomics, are unable to resolve the site-specific glycosylation pattern of glycoproteins.

The primary aim of this thesis was to develop glycoproteomic techniques for mass spectrometric analysis of glycoproteins. A sialic acid capture-and-release method, based on hydrazide chemistry, for selective enrichment of N- and O-linked glycopeptides from complex biological samples was developed. Enriched glycopeptides were separated by reversed phase liquid chromatography and analyzed by Fourier transform ion cyclotron mass spectrometry (FTICR MS) utilizing collision induced dissociation (CID) and electron capture dissociation (ECD) fragmentation techniques.

Initially, both N- and O-glycopeptides from sialylated glycoproteins of human cerebrospinal fluid (CSF) were enriched and characterized. Subsequently, a targeted O-glycoproteomics approach was developed, allowing for sequence analysis of preferred O-glycosylation sites of glycoproteins. The applicability of the sialic acid capture-and-release strategy was further demonstrated for human urine, a technically more challenging biological fluid. The LC-MS/MS analyses revealed unique N- and O-glycosylations, many of which were previously unknown, both for CSF and urinary glycoproteins. In e.g. CSF, a series of O-glycopeptides with Thr linked O-glycans in the vicinity of the β-secretase cleavage site of the amyloid precursor protein (APP) were identified. Additionally, amyloid beta (Aβ) peptides, originating from APP, were immunoprecipitated from CSF samples for a targeted glycoproteomic analysis. These analyses revealed that a series of Aβ peptides were uniquely modified with sialylated O-glycans at a specific Tyr residue. A relative increase of such Tyr O-glycosylated Aβ peptides was observed in CSF samples from Alzheimer's disease (AD) patients compared to non-AD patients, suggesting that these Aβ glycopeptides may potentially be used as biomarkers of AD.

Keywords: Glycobiology, Glycoproteomics, Mass spectrometry

ISBN: 978-91-628-8412-3

POPULÄRVETENSKAPLIG SAMMANFATTNING

Glykoproteiner är en särskild klass av proteiner som fyller många och skiftande funktioner i levande organismer. Glykoproteiner har olika typer av sockerstrukturer, även kallade glykaner, bundna till specifika aminosyror, vanligtvist till asparagin (N-glykaner) eller till serin/treonin (O-glykaner). Glykoproteinernas strukturer behöver kartläggas i detalj för att vi ska få en bredare förståelse av deras roller i fundamentala biologiska processer. Proteomikstudier har blivit ett väletablerat tillvägagångssätt för studier av alla samtidigt förekommande proteiner i komplexa biologiska blandningar och genom dessa kan man idag identifiera tusentals proteiner från enstaka prover. Många av dessa proteiner är glykoproteiner men information om de bundna glykanerna är ofta bristfällig och kräver särskild metodologi. Glykomik är en sådan kompletterande teknik för massspektrometrisk (MS) analys av glykaner. Specifik information om glykosylering erhålls vanligen då man, kemiskt eller enzymatiskt, separerar glykanerna ifrån proteinerna och studerar dessa sockerstrukturer separat. Dessa metoder ger god strukturell information om förekommande glykaner och proteiner men begränsas av att man inte kan ange exakt vilka enskilda glykaner som sitter bundna till vilka enskilda aminosyror på respektive protein.

Glykoproteomik är ett nytt forskningsfält där intakta glykopeptider, dvs. proteinfragment med kvarvarande glykaner, analyseras. Hittills har glykoproteomikstudier varit begränsade till analys av enskilda proteiner, mycket på grund av brist på effektiva anrikningsmetoder för glykoproteiner och glykopeptider. Det övergripande målet med denna avhandling har varit att utveckla nya metoder som möjliggör MS-baserade strukturstudier av glykopeptider avspjälkade från glykoproteiner anrikade ur komplexa biologiska prover.

Vår anrikningsstrategi baseras på en enkel kemisk modifiering av en speciell sockerstruktur, sialinsyra som är vanligt förekommande på många olika glykoproteiner, för att skapa reaktiva grupper som sedan kan utnyttjas för att koppla glykoproteinerna till kemiskt aktiverade plastkulor och därmed anrika dessa glykoproteiner på en fast yta. Genom att tvätta bort ospecifikt bundna proteiner kan de sialinsyreinnehållande glykoproteinerna anrikas selektivt. Efter klyvning av de bundna proteinerna t.ex. med enzymet trypsin, vilket generar många fragment av lämplig storlek för MS-analys, kan de kvarvarande glykopeptiderna selektivt frigöras från plastkulorna med mild syrabehandling. De frigjorda

glykopeptider, utan kvarvarande sialinsyror, kan därefter analyseras med MS.

Metoden tillämpades först på ryggmärgsvätska (CSF) och även om denna mängdmässigt domineras av albumin, som inte är ett glykoprotein, kunde vi effektivt anrika många olika glykoproteiner och karaktärisera deras N- och O-glykopeptider. Därefter förfinade vi tekniken för att möjliggöra riktad analys av enbart O-glykopeptider. Genom att inkludera en enzymatisk förbehandling, med PNGase F, som specifikt spjälkar alla glykaner kopplade till aminosyran asparagin kunde ännu fler O-glykopeptider från CSF anrikas, analyseras och identifieras. Anrikningsmetoden tillämpades framgångsrikt även på urin, som är ett tekniskt mer utmanande biologiskt utgångsmaterial. Metoden tillät oss att studera glykaner och deras kopplingar till olika aminosyror på enskilda anrikade glykopeptider, vilket inte är möjligt med tekniker som proteomik och glykomik. I dessa studier kunde vi beskriva flera tidigare okända glykosyleringar, framförallt O-glykosyleringar, och också dra slutsatser om vilka aminosyra-sekvenser som företrädesvis blir glykosylerade i den levande vävnaden (in vivo).

Många intressanta nya O-glykopeptider identifierades, bl.a. en glykopeptid, anrikad från CSF, som härstammar från amyloid prekursor proteinet (APP). Felaktig nedbrytning av APP i hjärnan ger en ökning av amyloid beta (Aβ) peptider som är giftiga för nervceller och som genom att aggregera bildar s.k. amyloida plack. Dessa amyloida plack tror man är avgörande för utvecklingen av Alzheimers sjukdom (AD). Med hjälp av antikroppar riktade emot APP kunde vi anrika APP/Aβ peptider ur CSF och med vår MS-analys påvisa att flera av dessa även var O-glykosylerade, vilket inte var känt sedan tidigare. Dessutom identifierade vi Aβ peptider med sialinsyreinnehållande O-glykaner bunda till tyrosin (Tyr), vilket inte setts tidigare i något mänskligt protein. En relativ ökning av dessa glykosylerade Aβ peptider observerades hos AD jämfört med icke-AD patienter vilket antyder att Tyr-glykosylerade Aβ peptider skulle kunna användas för tidig diagnostik av AD.

LIST OF PAPERS

This thesis is based on the following studies, referred to in the text by their Roman numerals.

I. Nilsson, J., Ruetschi, U., Halim, A., Hesse, C., Carlsohn, E., Brinkmalm, G., and Larson, G. (2009) Enrichment of glycopeptides for glycan structure and attachment site identification, *Nat Methods 6*, 809-811.

II. Halim, A*., Brinkmalm, G*., Ruetschi, U., Westman-Brinkmalm, A., Portelius, E., Zetterberg, H., Blennow, K., Larson, G., and Nilsson, J. (2011) Site-specific characterization of threonine, serine, and tyrosine glycosylations of amyloid precursor protein/amyloid beta-peptides in human cerebrospinal fluid, *Proc Natl Acad Sci U S A 108*, 11848-11853.

III. Halim, A., Nilsson, J., Ruetschi, U., Hesse, C., and Larson, G. (2011) Human urinary glycoproteomics; attachment site specific analysis of N-and O-linked glycosylations by CID and ECD, *Mol Cell Proteomics*. doi:10.1074/mcp.M111.013649

IV. Halim, A., Ruetschi, U., Larson, G., and Nilsson, J. Glycosylation motifs in the GalNAc O-glycosylation of cerebrospinal fluid proteins, *Manuscript*.

*Authors contributed equally to this work

Papers not included in this thesis:

Nilsson, J., Halim, A., Moslemi, A. R., Pedersen, A., Larson, G., and Oldfors, A. (2011) Molecular pathogenesis of a new glycogenosis caused by a glycogenin-1 mutation, *Biochim Biophys Acta*. http://dx.doi.org/10.1016/j.bbr.2011.03.031

Hesse, C., Johansson, I., Mattsson, N., Bremell, D., Andreasson, U., Halim, A., Anckarsater, R., Blennow, K., Anckarsater, H., Zetterberg, H., Larson, G., Hagberg, L., and Grahn, A. (2011) The N-terminal domain of alpha-dystroglycan, released as a 38 kDa protein, is increased in cerebrospinal fluid in patients with Lyme neuroborreliosis, *Biochem Biophys Res Commun 412*, 494-499

Karlsson, H., Halim, A., and Teneberg, S. (2010) Differentiation of glycosphingolipid-derived glycan structural isomers by liquid chromatography/mass spectrometry, *Glycobiology 20*, 1103-1116

Fagerberg, D., Angstrom, J., Halim, A., Hultberg, A., Rakhimova, L., Hammarstrom, L., Boren, T., and Teneberg, S. (2009) Novel Leb-like Helicobacter pylori-binding glycosphingolipid created by the expression of human alpha-1,3/4-fucosyltransferase in FVB/N mouse stomach, *Glycobiology 19*, 182-191.

Cederfur, C., Salomonsson, E., Nilsson, J., Halim, A., Oberg, C. T., Larson, G., Nilsson, U. J., and Leffler, H. (2008) Different affinity of galectins for human serum glycoproteins: galectin-3 binds many protease inhibitors and acute phase proteins, *Glycobiology 18*, 384-394.

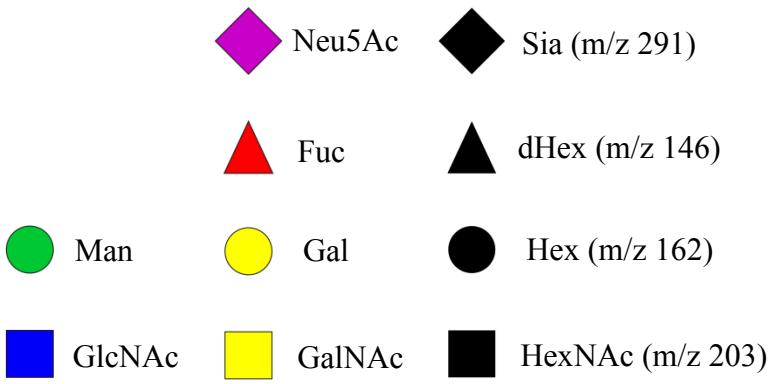
CONTENT

ABBREVIATIONS

AD	Alzheimer's disease
ADH	Antidiuretic hormone
APP	Amyloid precursor protein
Asn	Asparagine
Asp	Aspartic acid
BBB	Blood-brain barrier
CID	Collision induced dissociation
CMP	Cytidine-diphosphate
CNS	Central nervous system
CNX	Calnexin
CRT	Calreticulin
CSF	Cerebrospinal fluid
CTF	C-terminal fragment
Cys	Cysteine
dHex	Deoxyhexose
DNA	Deoxyribonucleic acid
ECD	Electron capture dissociation
EM	Electron multiplier
ER	Endoplasmic reticulum
ESI	Electrospray ionization
ETD	Electron transfer dissociation
FET	Bovine fetuin
FTICR	Fourier transform ion cyclotron resonance
Fuc	Fucose
Gal	Galactose
GalNAc	N-acetylgalactosamine
GC	Gas chromatography
GDP	Guanosine-diphosphate
Glc	Glucose
GlcA	Glucuronic acid
GlcNAc	N-acetylglucosamine
GPI	Glycosylphosphatidylinositol
Hex	Hexose
HexNAc	N-acetylhexosamine
HPLC	High performance liquid chromatography

IdoA	Iduronic acid
LacNAc	N-acetyllactosamine
LTQ	Linear ion trap quadrupole
m/z	Mass-to-charge ratio
MALDI	Matrix-assisted laser desorption/ionization
Man	Mannose
MS	Mass spectrometry
Neu5Ac	N-acetylneuraminic acid
NMR	Nuclear magnetic resonance
PC	Proprotein convertase
PNGase F	Peptide:N-glycosidase F
ppGalNAcT	Polypeptide-N-acetylgalactosaminyltrasferase
Pro	Proline
Ser	Serine
Sia	Sialic acid
SLe^X	Sialyl-$Lewis^X$
ST	Sialyltransferase
TF	Transferrin
Thr	Threonine
TOF	Time-of-flight
Tyr	Tyrosine
UDP	Uridine-diphosphate
Xyl	Xylose
ZFN	Zinc finger nuclease

Symbolic representations of monosaccharides



1 INTRODUCTION

1.1 Glycobiology

The principal building blocks in all cells are nucleic acids, lipids, proteins and carbohydrates [1]. In medicine, biochemistry and other related areas, carbohydrates are perhaps best known as molecules that store energy and fuel cellular metabolism. However, carbohydrates, or glycans, are also key components of various structural elements that are of central importance for cellular function and viability.

Remarkably, all living cells are covered with a dense layer of glycans [2]. This complex network, also known as the glycocalyx, is composed of free glycans and different types of glycoconjugates, i.e. oligosaccharides that are covalently attached to proteins or lipids. Oligosaccharides may exhibit an enormous structural complexity due to the variability of linkage positions and anomeric configurations that may connect epimeric monosaccharides in linear or branched sequences. Embedded in this complexity is a capacity to store biological information which, at least in theory, may exceed the storage capacity of other biopolymers, e.g. oligonucleotides and polypeptides [3]. It is therefore not surprising that every living organism has found a way to utilize glycans and glycan binding proteins to encode and decipher cellular information [4].

Although the basic questions of carbohydrate chemistry were elucidated very early, the modern understanding of glycan biosynthesis and molecular biology was not clarified until the second half of the 20th century. In fact, the specific term glycobiology surfaced for the first time in the late 1980s [5]. This scientific branch is engaged in the study of glycan- and glycoconjugate structures, glycan binding proteins (lectins), glycan biosynthesis, metabolism, evolution and biology in health and disease.

1.1.1 Glycans

Basic structural units of glycans

Monosaccharides are aldehydes or ketones with multiple hydroxyl groups. One of the simplest monosaccharides is glyceraldehyde with three carbon atoms, referred to as *triose*, and two hydroxyl groups. The general formula $(C\text{-}H_2O)n$ is applicable for most monosaccharides and literally means "hydrated carbon". Monosaccharides with six carbon atoms, or *hexoses*, are

typically found as constituents of glycoconjugates. Although the connecting order of atoms in a monosaccharide may be the same, the geometrical orientation of hydroxyl groups might be different, giving rise to stereoisomers, e.g. glucose (Glc), galactose (Gal) and mannose (Man) [6]. These hexoses, together with their N-acetamido derivatives N-acetylgalactosamine (GalNAc) and N-acetylglucosamine (GlcNAc), are some of the basic building blocks of vertebrate glycans. A special family of negatively charged, 9 carbon monosaccharides, termed N-acetylneuraminic acid (Neu5Ac) or just sialic acid (Sia), are typically found at terminal positions of glycans (see below) [7, 8]. Additional constituents of glycans include fucose (Fuc), xylose (Xyl), glucuronic acid (GlcA) and iduronic acid (IdoA) [9].

Monosaccharides have a tendency to form cyclic structures in solution. This process involves the reaction of a hydroxyl group with the C-1 aldehyde or ketone, resulting in the formation of a hemiacetal group. Any of the anomeric forms, designated α and β, may result from this reaction (Fig. 1) [10].

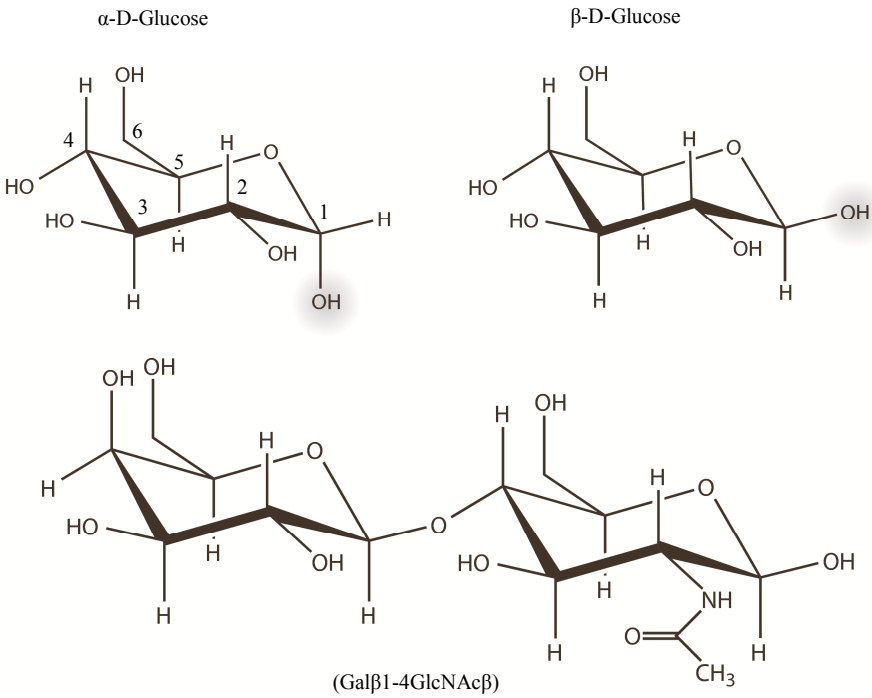


Figure 1. The D-glucose hemiacetals are shown as α- and β anomers (top). The structure of N-acetyllactosamine (LacNAc) is shown (bottom).

The anomeric carbon of one monosaccharide may be joined with the hydroxyl group of another. This reaction (enzymatically catalyzed, see below) forms the O-glycosidic bond that links monomeric units into oligosaccharides. Exemplified in Fig. 1, the anomeric C-1 of Gal has been combined with the C-4 hydroxyl group of GlcNAc, resulting in the formation of typical disaccharide found in glycans, namely Galβ1-4GlcNAc (LacNAc). According to this principle, stereoisomers may be linked as α- or β anomers at different positions to form numerous linear or branched glycans. Each structure is considered to be a unique epitope with a potentially distinct biological activity.

Glycan biosynthesis

The primary synthesis of glycans is orchestrated by the endoplasmic reticulum (ER) and Golgi associated glycosyltransferases. These enzymes catalyze the transfer of activated monosaccharides, or nucleotide sugar donors, to acceptor substrates, e.g. lipids, proteins or other glycans. Most monosaccharides are activated as uridine-diphosphate (UDP) sugars, e.g. UDP-Glc, UDP-Gal, and UDP-GalNAc. The only monosaccharides that deviate from this rule are Fuc, Man and Neu5Ac, which are found as guanosine-diphosphate (GDP-Fuc and GDP-Man) or as cytidine-monophosphate (CPM-Neu5Ac) sugars [9]. Each nucleotide donor sugar is recognized by a specific family of glycosyltransferases, of which each family member is able to catalyze a specific reaction between donor and acceptor substrates. Glycan biosynthesis can therefore be generalized according to the "one enzyme-one linkage" hypothesis, i.e. a specific nucleotide donor sugar is linked to a specific acceptor substrate in one particular anomeric configuration and linkage position [11]. Following synthesis, glycans may also be trimmed (by various glycosidases) or further modified by e.g. phosphorylation, sulfation or acetylation [12-15]. Collectively, these processes are able to produce highly complex glycan structures [16].

Glycoconjugates

Molecules composed of glycans that are covalently and enzymatically attached to non-carbohydrate units are generally classified as glycoconjugates. Glycoconjugates are widely distributed in biological fluids, cellular membranes and intracellular compartments. This large family of molecules includes e.g. glycoproteins, proteoglycans, glycosphingolipids, glycosylphosphatidylinositol (GPI) anchored proteins and glucuronides. However, the discussion will from here on only focus on the concepts of glycoproteins, their biosynthesis, structures and functions.

1.1.2 Glycoproteins

N-linked glycoproteins

Glycoproteins are usually categorized according to the nature of the linkage by which the glycans are attached. For N-linked glycoproteins, or simply just N-glycoproteins, the oligosaccharide component is linked to the side chain amide nitrogen (N) atom of asparagine residues (Asn), hence the names N-glycan and N-glycoprotein. Although other forms exist, GlcNAcβ1-*N*-Asn represents the most common type of N-glycosidic linkage to proteins [17].

N-glycan biosynthesis is initiated in the ER by the *en bloc* transfer of a preassembled $Glc_3Man_9GlcNAc_2$ moiety from a dolichol-pyrophosphate lipid anchor to newly synthesized proteins. The asparagine residue can accept the $Glc_3Man_9GlcNAc_2$ moiety only if it is part of an asparagine-X-serine/threonine/cysteine consensus sequence (Asn-X-Ser/Thr/Cys), where X is any amino acid except proline (Pro) [18, 19]. Following transfer to an Asn residue, the oligosaccharide is processed by specific glycosidases that hydrolyze individual glycosidic bonds of the 14-sugar glycan. The monoglucosylated $Glc_1Man_9GlcNAc_2$ intermediate, which is recognized by the lectin-like ER chaperones calnexin (CNX) and calreticulin (CRT), plays a central role in protein quality control and folding [20, 21]. Properly folded N-glycoproteins are subsequently translocated to the Golgi compartments and processed by various glycosidases and glycosyltransferases to produce one of three different types of N-glycan structures, i.e. complex, hybrid or high-mannose.

Each type consists of a core $Man_3GlcNAc_2$ glycan with at least two branched glycan chains, also known as antennas. The three different N-glycan types are classified according to the sugar composition of the antennae. Complex type N-glycan antennas are composed of different monosaccharides (Fig. 2), e.g. Neu5Acα2-6Galβ1-4GlcNAcβ2Man, whereas high-mannose type N-glycans contain only Man units on both antennas. Hybrid type N-glycans are a combination of high-mannose and complex N-glycans [22].

The final step of maturation occurs in the late Golgi compartments where most N-linked oligosaccharides are converted to complex type N-glycans before the glycoprotein is secreted or translocated to the plasma membrane. In humans, complex type N-glycans constitute the dominating glycoform of secreted N-glycoproteins [23].

The complex type N-glycans represent a heterogeneous mixture of structures that accommodate not only other monosaccharides (e.g. Fuc or GalNAc) but also include sulfate or phosphate modifications. The structural diversity of these glycans is further extended by the addition of three (tri-antennary) or four (tetra-antennary) antennas with terminal Neu5Ac units. Consequently, the population of a single glycoprotein may comprise several unique N-glycan structures, situated at one or multiple Asn-X-Ser/Thr/Cys N-glycosylation sites. The number of complex type N-glycans can therefore exceed 150 different structures for individual glycoproteins [24].

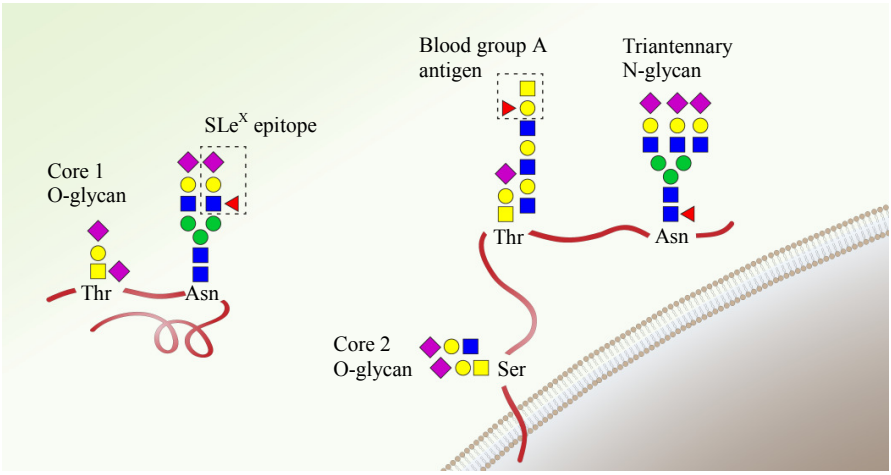


Figure 2. Secreted and membrane bound glycoproteins.

O-linked glycoproteins

The covalent attachment of glycans to the hydroxyl groups of Ser and Thr residues constitutes the structural basis of O-linked glycoproteins (O-glycoproteins). Protein linked O-glycans can be initiated through many different monosaccharides but the most abundant and complex form of O-glycosylation is generally attributed to the mucin-type GalNAcα-O-Ser/Thr glycans [25]. In contrast to N-glycans, mucin-type O-glycans lack a common core structure and are all elongated from a single monosaccharide to form eight different core structures (table 1) [26]. The simplest mucin-type glycan, sometimes referred to as the Tn antigen, is thus GalNAcα-O-Ser/Thr. This single monosaccharide is transferred from UDP-GalNAc to Ser/Thr residues by the action of polypeptide-N-acetylgalactosaminyltransferases (ppGalNAcT) [27].

Table 1. Core structures of mucin-type O-glycans.

O-Glycan	Structure
Tn-antigen	GalNAcα-O-Ser/Thr
Core 1 (T antigen)	Galβ1-3GalNAcα-O-Ser/Thr
Core 2	GlcNAcβ1-6(Galβ1-3)GalNAcα-O-Ser/Thr
Core 3	GlcNAcβ1-3GalNAcα-O-Ser/Thr
Core 4	GlcNAcβ1-6(GlcNAcβ1-3)GalNAcα-O-Ser/Thr
Core 5	GalNAcα1-3GalNAcα-O-Ser/Thr
Core 6	GlcNAcβ1-6GalNAcα-O-Ser/Thr
Core 7	GalNAcα1-6GalNAcα-O-Ser/Thr
Core 8	Galα1-3GalNAcα-O-Ser/Thr

The enzyme family consists of at least 20 members, ppGalNAcT-1 to -20, that differ in the ability to recognize and glycosylate their polypeptide substrates [25, 28]. Contrary to N-glycosylation, mucin-type O-glycosylation lacks a consensus peptide sequence although these O-glycans are generally found in protein regions rich in Ser, Thr and Pro residues [29].

In the majority of tissues, the second step of mucin-type O-glycan synthesis involves the addition of Gal to the Tn antigen, resulting in the formation of core 1 O-glycans (T antigen) [30]. This specific step is performed by T-synthase and requires the molecular chaperone Cosmc for activity in the Golgi compartment [29, 31]. The core 1 O-glycan biosynthesis is usually terminated by the addition of two Neu5Ac units at C-3 of Gal and C-6 of GalNAc. This reaction may be catalyzed by several enzymes of the sialyltransferase (ST) family, e.g. ST3Gal I and ST6GalNAc I members [29]. The outcome of this pathway results in the formation of di-sialylated core 1 O-glycans (Fig. 3).

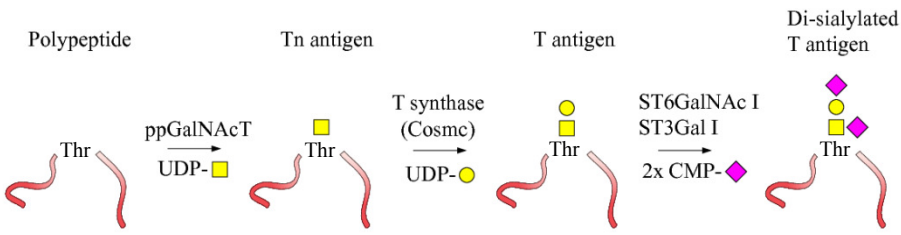


Figure 3. Biosynthesis of di-sialylated core 1 O-glycan.

Alternatively, the Tn and T antigens may be modified by GlcNAc prior to the addition of Neu5Ac to form core structures 2, 3 or 4. Core 2 O-glycans are found in most tissues whereas core 3 and core 4 O-glycans are prominent structures on the mucus epithelia of the gastrointestinal and respiratory tracts [29, 32-36]. Core structures 5-8 are rare and the enzymatic processes that generate these O-glycans are essentially unknown [25].

Core structures 2-4 are typically elongated to complex O-glycan structures by the addition of repeating disaccharides, e.g. poly-LacNAc (Fig. 1), or other biologically active epitopes. The synthesis of complex O-glycans is usually terminated by the addition of histo-blood group antigens or terminal sialic acids (Fig. 2) [29].

Biological roles of N- and O-linked glycans

Although N- and O-linked glycans have different core structures and protein attachment sites, the true structural diversity and biological activity is often found at terminal positions of these glycans. Secreted and membrane bound N- and O-glycoproteins are typically decorated with Neu5Acα2-3Galβ-R or Neu5Acα2-6Galβ-R epitopes that are utilized as ligands by various endogenous lectins. These interactions include intrinsic recognitions by Siglecs (I-type lectins) in e.g. cell-cell communications and signaling events of the immune, haemopoietic and nervous systems [37, 38]. Terminal sialic acids are also important constituents of larger epitopes, e.g. the Neu5Acα2-3Galβ1-4(Fucα1-3)GlcNAcβ-R glycan, also known as the sialyl-Lewis^X (SLe^X) determinant (Fig. 2). Together with its binding partners E-, L- and P-selectins, the SLe^X epitope participates in leukocyte-endothelial cell adhesions and plays a key role in the inflammatory response process [39, 40].

Heterogeneous structures of branched and elongated mucin-type O-glycans are usually found in Pro, Ser and Thr rich protein domains of secreted and membrane bound mucins. These clusters of hydrophilic and negatively charged O-glycans are able to bind H_2O and form a viscous coat that lubricates and protects the underlying tissue against chemical and microbial insult [25, 29, 41, 42]. Mounting evidence also suggests that single, sialylated core 1 O-glycans located at specific Ser/Thr residues adjacent to enzymatic cleavage sites are able to suppress proteolytic processing of proteins [26, 28, 43-45]. Proprotein convertase (PC) processing, responsible for producing active neuronal and endocrine hormones by regulated proteolysis of secretory proteins [46], has also been shown to be co-regulated by site-specific O-glycosylations [47-49].

The ABO(H) blood group system is perhaps the most illustrative example of a specific biological consequence of glycans. These determinants are composed of a common Fucα1-2Galβ-R structure, i.e. the H antigen, and are typically found as terminal epitopes on glycosphingolipids and N- and O-linked glycans (Fig. 2) [16, 50-52]. Depending on the genetic background of the individual, the terminal Fucα1-2Galβ-R disaccharide may be branched as GalNAcα1-3(Fucα1-2)Galβ-R (A antigen), as Galα1-3(Fucα1-2)Galβ-R (B antigen) or remain unmodified (H antigen). These structures are of central importance for inter-individual blood transfusions and organ transplantation procedures.

However, the glycans described above represent only a few, selected examples of biologically active epitopes. Many other functional roles have been assigned to protein linked glycans and the biological significance of some oligosaccharides are still unknown [53]. Akin to "junk" DNA, there are presently also "junk" glycans with no apparent biological activity [54]. Depending on type, the biological role of a given glycan may thus be vital or, as yet, not fully understood.

1.2 Mass spectrometry

Mass spectrometry (MS) is one of the most important analytical techniques in structural glycobiology [55]. The principles of MS are based on ionization of molecules, separation of gas-phase ions and detection of these charged species according to their mass-to-charge ratio (m/z). Thus, in its most basic form, MS can be used to determine the mass of a given analyte. However, modern MS instruments are also able to dissociate ionized molecules (precursor ions) and mass measure the resulting fragments (product ions). This process, referred to as tandem MS, MS/MS or MS^2, can therefore be used to elucidate the primary structure (sequence) of peptides, glycans or glycopeptides.

Although the design and configuration may vary considerably, all MS instruments are composed of three primary components, i.e. an ion source, a mass analyzer and a detector. The ion source generates gas-phase ions, an absolute requirement in MS, from nonvolatile, thermally labile analytes, e.g. proteins or peptides. The most common ionization mechanisms of modern mass spectrometers are matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI) [56, 57]. The characteristic behavior of charged particles in electric/magnetic fields allows the mass analyzer to separate specific ions according to their m/z ratio. Many modern MS instruments have e.g. time-of-flight (TOF), linear ion trap quadrupole (LTQ),

Fourier transform ion cyclotron resonance (FTICR) or orbitrap mass analyzers. The ion detection process is usually amplified by a separate electron multiplier (EM) or recorded as an image-current (frequency) by the mass analyzer itself (for FTICR and orbitrap) [58]. Regardless of instrumental design and configuration, the MS data is always presented as a mass spectrum, i.e. a plot of ion intensities (y-axis) vs. ion m/z values (x-axis).

In addition, two separate mass analyzers, e.g. LTQ and FTICR, may be combined in the same instrument for optimal performance. The resulting hybrid mass spectrometer may also be interfaced with a high performance liquid chromatography (HPLC) system to facilitate the analysis of complex biological samples. In the remaining part of this thesis, the principles and applications of ESI-LTQ-FTICR MS, a hybrid mass spectrometer used to collect the MS data of this thesis, will be described in more detail.

1.2.1 ESI-LTQ-FTICR mass spectrometry

Electrospray ionization

The underlying theory for dissipating a liquid sample as a homogeneous electrospray was formulated during the 1960s by Sir Geoffrey Taylor and, shortly after, Dole and coworkers speculated that electrospray could be used as a technique to generate molecular beams of larger molecules [59, 60]. However, it was not until the late 1980s that ESI of proteins was shown to be possible [57, 61].

ESI is initiated when a potential is applied to the edge of a capillary. The electrostatic force pulls the liquid towards the tip of the capillary and causes it to adopt an elliptic shape. At a certain voltage, the curvature of the ellipse collapses into a cone, also known as a Taylor cone, resulting in the formation of a spray that starts to emit fine droplets of liquid [62]. Depending on the polarity of the applied potential, the droplets may contain either positively or negative charged molecules. While in flight, the droplets shrink in size by evaporation of neutral solvent particles [63], a process that proceeds until multiply charged molecular ions are formed at atmospheric pressure [64]. Although the exact mechanism is still a matter of debate, two models have emerged to explain the details of ESI, namely the ion evaporation model [65, 66] and the charge residue model [60, 62].

ESI offers many advantages to MS-based analysis of biomolecules. Because it ionizes molecules directly from the liquid phase, the ESI ion source may be coupled to an HPLC system to produce gas-phase ions during the

chromatographic separation (on-line HPLC-MS) of e.g. peptides or glycans. ESI is also considered to be a "soft ionization" technique able to generate molecular ions without compromising the integrity of large molecules or protein complexes [67]. These properties, together with the high efficiency (up to 100%) of ESI [62, 68], have made this ionization method a popular choice in many MS-based analyses.

LTQ-FTICR mass spectrometry

Comisarow and Marshall developed the first FTICR mass analyzer [69, 70] based on the original ICR MS technique [71]. This method is based on the fact that charged particles are manipulated in magnetic fields. More specifically, the velocity vector of a charged particle experiences a force, referred to as the Lorentz force, when subjected to a perpendicular magnetic field. The Lorentz force, which is perpendicular to the plane defined by the magnetic field and the velocity vector, therefore causes the charged particle to adopt a circular trajectory [72]. Since the magnetic field (B) is constant and uniform in FTICR MS, the periodic rotation of the charged particle, which can be expressed as a frequency (f), will thus only depend on the mass (m) and charge (z) of the particle, according to the equation:

$$f = zB/2\pi m$$

Given that frequencies can be measured very accurately, a mathematical Fourier transform operation may therefore be used to convert the frequency of each analyte ion to an accurate mass. The hallmarks of FTICR MS are therefore high mass accuracy and high mass resolution [73-75].

Although the FTICR mass analyzer is capable of performing tandem MS experiments, there is a significant advantage when analyte ions are fragmented in a linear ion trap. The LTQ is characterized by e.g. fast scanning ability and excellent ion storage capacity which improves the overall sensitivity for product ion detection during tandem MS analysis [76]. The hybrid LTQ-FTICR mass spectrometer is therefore able to perform accurate mass measurements of precursor ions and sensitive detection of product ions [77]. The fragmentation mechanisms of the LTQ and FTICR mass analyzers are further described in the methods section.

1.2.2 MS applications and –omics

Proteomics

Proteomics is an MS-based technology that generally deals with the identification and characterization of proteins. More specifically, proteomics

is defined as the qualitative or quantitative study of the protein expression profile in a specific organism, tissue or cell during a specific physiological state [78, 79]. A typical proteomics experiment includes a pre-analytical step where proteins are initially reduced and alkylated (Cys residues). Subsequently, the proteins are digested into peptides by specific proteases, e.g. trypsin (cleaves C-terminally to arginine/lysine-X, where X ≠ proline). The tryptic peptides are passed through a HPLC system, separated according to their physiochemical properties and eluted directly into the ESI ion source. The on-line HPLC approach allows a subset of peptides to be analyzed during each time interval of the chromatography, thereby increasing the overall sensitivity of the analysis [80]. At a given retention time, peptides are thus ionized, mass measured and fragmented by the mass spectrometer. The experimental dataset is subsequently identified by probability based scoring algorithms, e.g. Mascot [81]. In MS jargon, this is usually referred to as a "database search" or a "Mascot search" since it involves the automated matching of experimental data to theoretical peptide sequence data stored in a public database, e.g. Swiss-Prot or NCBInr. This type of proteomics experiment, also known as bottom-up proteomics [82], is a very powerful tool for identification of proteins in complex biological samples [83].

In addition, various pre-analytical strategies may be applied to the common proteomics workflow for maximal analytical sensitivity. These include e.g. analysis of specific 1D or 2D SDS-PAGE bands/spots for characterization of individual proteins [84-86], enrichment of phosphorylated proteins for targeted phosphoproteomics [87, 88] or lectin affinity enrichment for identification of glycoproteins [89, 90].

However, it should be stressed that the probability based scoring algorithms are primarily designed for identification of peptides, which are subsequently grouped into protein identities. Although simple modifications, e.g. phosphorylations, are readily identified in a database search [83], the probability based scoring algorithms are unable to resolve the structural complexity of glycans. Thus, even though glycoproteins *per se* may be identified, the protein linked glycans are not characterized by the common proteomics approach.

Glycomics

Glycomics is the systematic study of the glycan repertoire, i.e. the glycome, synthesized by a cell, tissue or organism under specific conditions. Although other technologies are available, the primary tool for analyzing the glycome is MS [91]. Most commonly, glycans are separated from their lipid- or protein components to allow high-throughput analysis. Protein linked O-

glycans are usually released as alditols by reductive beta elimination [92, 93] whereas N-glycans are released by enzymatic peptide:N-glycosidase F (PNGase F) treatment [94-96]. Prior to MS analysis, released glycans may also be derivatized or labeled with various reagents to improve chromatographic separation or structural identification [97].

A complete structural characterization of any glycan includes determining the size of the glycan, the identification of all monosaccharides included, their sequence (and possible branching), their relative linkage positions and determination of anomeric configurations of all the glycosidic bonds. For released glycans, reductive amination may be employed to couple fluorophores to glycans and thereby facilitate the quantification of the released glycoforms in each sample. Alternatively, gas chromatography-mass spectrometry (GC-MS) may be employed to obtain quantitative information on the molar ratios of individual monosaccharides in a sample. However, this technique requires the glycans to be hydrolyzed into monosaccharides prior to analysis. In addition, the hydrogen atoms of hydroxyl-, carboxyl- and amine groups may be replaced by methyl groups in a reaction known as permethylation. Tandem MS analysis of the permethylated glycans may yield detailed information on the linkage positions in individual glycans. Collectively, these techniques may thus yield qualitative and quantitative information on glycan composition and linkage positions. However, the anomeric configuration is not easily addressed by MS-based methods. One useful approach is to employ exoglycosidases to hydrolyze specific glycosidic bonds in a step by step manner. The digested glycans may subsequently be analyzed by GC-/LC-MS based methods. Alternatively, nuclear magnetic resonance (NMR) may be employed in order to characterize the absolute glycan structure but this technique requires large amounts of pure glycans and is thus not as sensitive as MS-based methods.

Glycoproteomics

In glycoproteomics, MS-based methods are used to elucidate not only glycoprotein identities but also to characterize glycan structures and pinpoint their protein attachment sites. This methodology is thus distinguished by the absolute requirement of intact glycopeptides for analysis. The combination of complex glycan structures on variable peptide sequences renders the glycoproteome orders of magnitude more complex than e.g. the glycome [98]. Consequently, glycoproteomic analysis may be considered even more challenging by comparison to proteomics or glycomics.

Efficient enrichment of glycopeptides is a prerequisite in any glycoproteomic experiment. The primary pre-analytical task is thus elimination of non-

glycosylated proteins. Equally important, the non-glycosylated tryptic peptides originating from glycoproteins must be removed to avoid analytical interference. Several different approaches have been used for this purpose [99, 100], including lectin affinity purification [101], size-exclusion [102], and hydrophilic interaction liquid chromatography [103, 104]. However, these methods have generally been restricted to the studies of single glycoproteins and their applicability for glycoproteomic characterization of complex biological samples has not been shown.

The identification of glycopeptides, i.e. their glycans, attachment sites and peptide sequences, is far from simple, even when a pure glycopeptide sample is analyzed. Algorithms or databases are not available for automated identifications, thereby leaving no other option but to annotate glycopeptide spectra manually [100]. Mucin-type O-glycopeptides are particularly difficult to analyze. In contrast to N-glycosylation, O-glycan attachment sites cannot be predicted by the primary amino acid sequence and require complementary fragmentation modes for identification. In addition, variable O-glycosylation, with respect to O-glycan density and structural heterogeneity, makes O-glycopeptides exceptionally difficult to characterize [105, 106]. It is therefore not surprising that only a few glycoproteomic studies, aimed at analyzing glycopeptides from complex biological samples, have been published so far [107-109].

Nevertheless, glycoproteomic characterization of biological samples is becoming increasingly important [110]. Although the combination of proteomics and glycomics is a powerful approach to study glycoproteins, a comprehensive understanding of glycoprotein structure and function also requires a detailed knowledge of the site-specific location of glycan structures. Given that aberrant glycosylation is associated with various pathological conditions [111-113], glycoproteomic investigation may also prove valuable for the discovery of novel biomarkers. This technique is therefore of central importance for further examination of glycoproteins in relation to health and disease. The section below provides a more detailed discussion on the biological fluids that we have used to explore the human glycoproteome.

1.3 Cerebrospinal fluid

Cerebrospinal fluid (CSF) is a clear and colorless fluid that surrounds the brain and the spinal cord. Although CSF is 99% water, it still contains a complex mixture of proteins, peptides and micronutrients (vitamin C and B6, folates etc.) [114]. In many ways, the composition of CSF is similar to blood,

with somewhat less glucose content and slightly higher concentration of NaCl. However, CSF is distinguished by the lack of cellular elements and relatively low protein content (~0.35 g/L) in comparison to blood.

The majority of CSF is produced by the *choroid plexuses* (Fig. 4) through a dynamic process involving passive filtration, pinocytosis and active secretion [115, 116]. Newly produced CSF exits the *third* and *fourth ventricle* through three openings (median and lateral apertures) and enters the *subarachnoid space* to reach the *cerebellomedullary*- and *pontine cisterns*. Subsequently, CSF flows at about 0.35 mL/min through the *subarachnoid space* and over the surface of the brain towards the area of *superior sagittal sinus*. Most of the CSF is then absorbed back into the venous system at the point of *arachnoid granulations* [116]. The ventricular system and *subarachnoid space* contains approximately 140 mL CSF which is replaced three to four times each day. In other words, approximately 500 mL CSF is produced by the *choroid plexuses* during a 24 hour period.

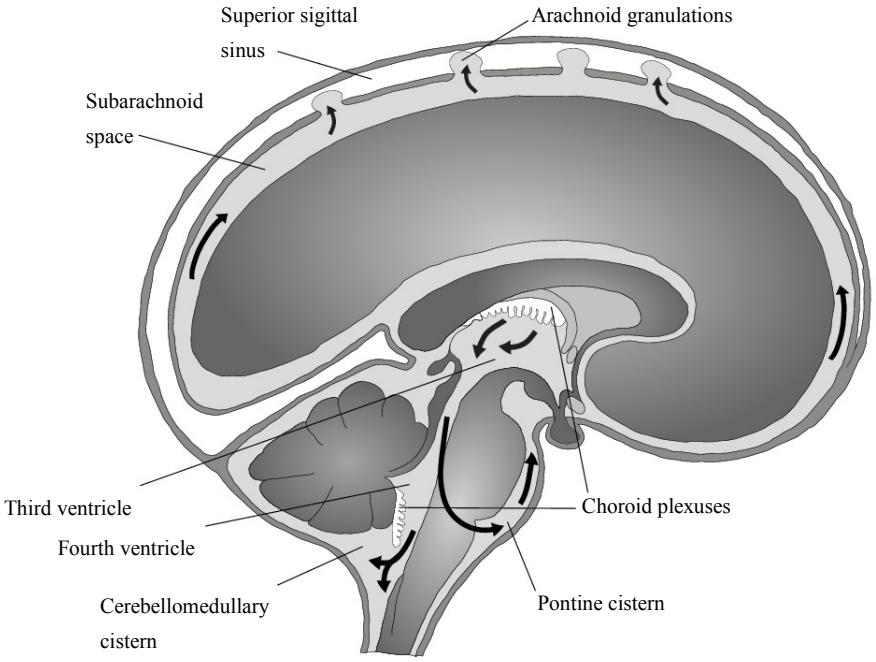


Figure 4. Cerebrospinal fluid excretion, circulation and reabsorption.

One basic function of CSF is to provide mechanical support by reducing (>30-fold) the effective weight of the brain. The buoyancy of this fluid also helps to absorb the shearing forces of acceleration generated during rapid head movements. Another important function of CSF is to transport macromolecules, e.g. transthyretin, insulin-like growth factor and thyroxine, to target neuronal cells and maintain a constant supply of micronutrients in circulation. Analogously, harmful metabolites are transported by CSF to the area of *arachnoid granulations* where they are absorbed back into the venous system. Also, CSF maintains the intracranial pressure in response to blood volume and retains a constant buffer reservoir of osmolytes which can be exchanged with the brain interstitial fluid to counteract increments in ion concentrations. In essence, CSF provides both a physical and a biochemical environment that is suitable for the brain and the spinal cord.

The main fraction of CSF proteins is derived from serum by the action of pinocytosis across the capillary endothelium. Albumin alone constitutes 55-75% of the total protein content in CSF, which, together with certain immunoglobulins (>15%), accounts for the dominating portion of the CSF proteome by mass [117]. However, secretory mechanisms of the *choroid plexuses* also release proteins that can be regarded as CSF-specific. The close proximity of CSF to the surrounding brain tissue allows products of protein catabolism or tissue degeneration to be regularly released into the CSF, resulting in a characteristic protein profile for CSF. Qualitatively, it is estimated that 56% of the proteome is CSF-specific since these proteins are not detected in the plasma proteome [118].

Pathological conditions in the central nervous system (CNS) may alter the composition of CSF and cause changes in CSF protein profiles. Due to the close contact of CSF to the surrounding brain tissue, neuropathological conditions are more likely to be reflected in the CSF proteome than in other biological fluids, e.g. plasma or urine. CSF therefore provides a natural source for biomarkers that can be useful for diagnosing, predicting or monitoring the progression of diseases in the CNS.

1.3.1 Alzheimer's disease and APP

Alzheimer's disease (AD), first described by the Bavarian psychiatrist Alois Alzheimer in 1907 [119], is the most common form of age-related dementia. Approximately 20-30 million people worldwide are believed to be affected by dementia, of which AD accounts for 50-60% of all cases [120].

AD is a progressive neurodegenerative disorder of the CNS which results in neuronal and synaptic loss in the cortical areas of the brain. The initial symptoms of mild cognitive decline, disorientation, impaired judgment and short term memory loss gradually intensify and ultimately results in dementia, which clinically defines AD. Neuropathological features of the AD-affected brain include extracellular deposition of amyloid plaques and intraneuronal neurofibrillary tangles. The latter are formed by hyperphosphorylation of the microtubule-associated protein tau, which induces self-assembly of tangles [121]. The extracellular amyloid plaques are mainly composed of aggregated amyloid beta (Aβ) peptides [122-124] which are derived from the amyloid precursor protein (APP) [125, 126].

APP is a ubiquitously expressed type 1 transmembrane glycoprotein for which the precise physiological roles remain unknown. The catabolic Aβ peptides result from sequential proteolytic processing of APP by the action of

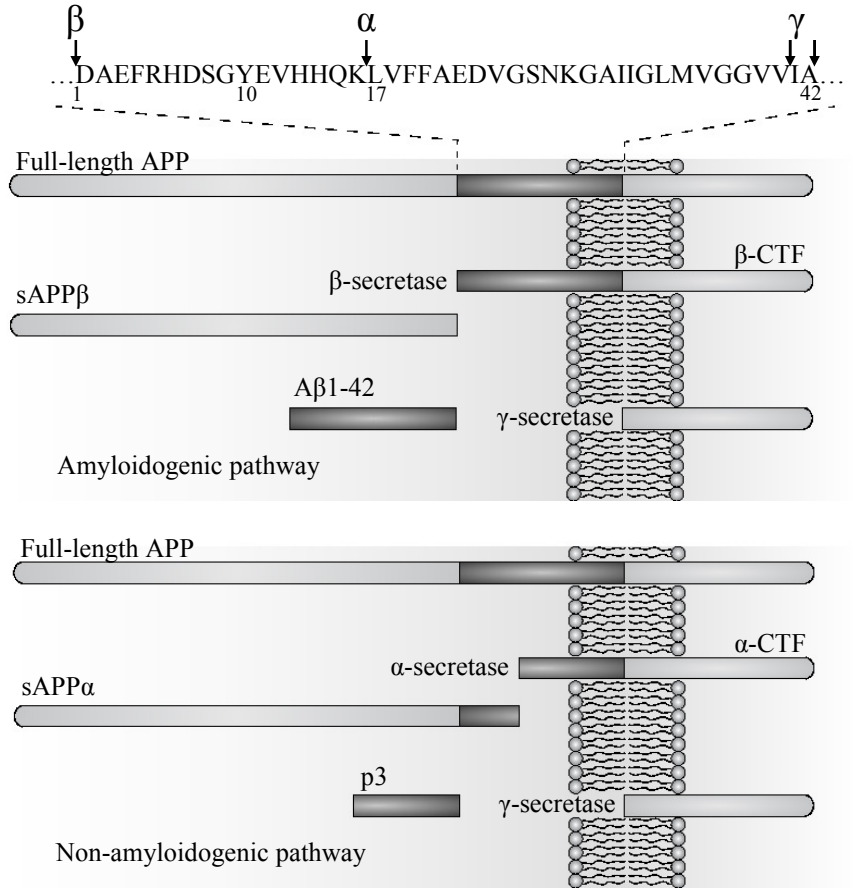


Figure 5. Proteolytic processing of APP by α-, β- and γ-secretases

α-, β- and γ-secretases (Fig. 5) [127-129]. In the amyloidogenic pathway, full-length APP is initially cleaved by β-secretase to release a soluble APP ectodomain (sAPPβ) while leaving a C-terminally truncated fragment (β-CTF) in the membrane. The β-CTF is subsequently processed by intramembranous γ-secretase cleavage which liberates the 40-42 amino acid peptides (Aβ1-40 and Aβ1-42). The Aβ1-40 isoform is quantitatively more abundant in CSF but the two additional amino acids of Aβ1-42 enhance the hydrophobic character and amyloidogenic properties of this isoform. Aβ1-42 is the major constituent in amyloid plaques but both variants can be detected in these aggregates. In the non-amyloidogenic pathway, full-length APP is cleaved by α-secretase to release the sAPPα ectodomain. The α-secretase cleavage occurs at amino acid 16 (numbering according to the Aβ sequence) and thereby precludes the formation of e.g. Aβ1-42. Instead, the following γ-secretase cleavage results in formation of the p3 fragment (Aβ17-40/42) which is believed to be released into CSF.

In addition to the processing pathways described above, APP/Aβ can undergo further proteolytic cleavages to yield alternative Aβ fragments in CSF [130-132]. The β-CTF can be cleaved by α-secretase to give shorter Aβ1-X fragments (e.g. Aβ1-14/15) whereas alternative γ-secretase cleavage of β-CTF can yield e.g. Aβ1-17/18/19/34/38 variants [133]. Cleavage of APP by the joint activity of α-secretase and other (unidentified) proteases can also produce endogenous APP/Aβ fragments that are N-terminally truncated upstream of the β-secretase cleavage site. These APP/Aβ fragments are typically cleaved at residues -57, -51 or -25 (relative to Asp1 of the Aβ sequence) and extend to Gln15 of the Aβ sequence [134]. Collectively, these are referred to as APP/AβX-15 fragments.

Aβ peptides are constitutively produced at high levels under non-pathological conditions in humans. This process is normally counterbalanced by clearance mechanisms that eliminate Aβ species, e.g. Aβ1-42, from CSF [120]. These mechanisms include receptor-mediated transport across the blood-brain barrier (BBB) and proteolytic degradation of Aβ peptides by e.g. neprilysin (NEP).

Twenty years ago, the amyloid cascade hypothesis was proposed as the initiating event in AD [135, 136]. According to this hypothesis, the intricate balance between Aβ production and clearance is disrupted, leading to the accumulation and aggregation of Aβ peptides (mainly Aβ1-42) which subsequently deposit as amyloid plaques. This process is believed to trigger the neuropathological changes, i.e. neuronal and synaptic loss, which ultimately results in cognitive dysfunction and dementia.

1.4 Urine

Urine production by the kidneys produces a transparent, colorless to dark yellow fluid that contains a broad class of substances, e.g. electrolytes (salts), vitamins, amino acids and other organic compounds. Proteins and peptides are present but in significantly lower quantities by comparison to plasma or CSF. For healthy individuals, protein content in urine does usually not exceed 0.1 g/L. Normal cellular activity, e.g. protein or intermediary metabolism, generates various metabolic end products that are constitutively released into the bloodstream. One important function of the kidneys is to prevent toxic build-up of such metabolic waste products through the process of urinary elimination.

Twenty-five percent of the cardiac output, i.e. ~2000 L/day, is circulated and filtered in the kidneys. This process produces the primary urine, or glomerular filtrate (~180 L/day), of which 99% is reabsorbed in the kidneys before the final urine (~1.5 L/day) is excreted to the bladder [137].

The formation of urine is precisely regulated by three distinct processes, i.e. glomerular filtration, tubular reabsorption and tubular secretion [138-140]. These processes occur in the nephrons (Fig. 6), which are the functional units of the kidney. Urine formation begins with the filtration of plasma at the

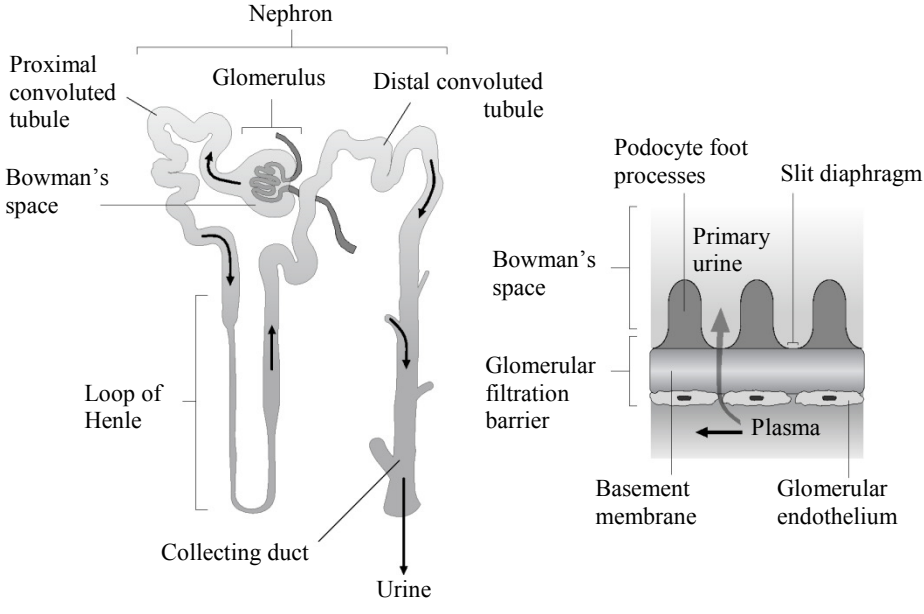


Figure 6. Production of urine by the process of glomerular filtration, tubular reabsorption and secretion in the nephron.

glomerulus. The pores of the *glomerular endothelium* together with the *basement membrane* and the *slit diaphragm* structure form a semipermeable barrier in three levels (*glomerular filtration barrier*) that allows free passage of small molecules, e.g. H_2O and electrolytes [140]. However, the passive filtration of large molecules (>60 kDa) and proteins, especially negatively charged proteins, is efficiently restricted by this barrier [141]. The glomerular filtrate subsequently flows from *Bowman's space* into the *proximal convoluted tubule* where the majority of H_2O, electrolytes, glucose, amino acids, peptides and proteins are reabsorbed. Organic acids, hydrogen ions and ammonia are simultaneously secreted into the fluid by the tubular cells. This dynamic process proceeds in each substructure of the *loop of Henle*, where a verity of substances are constitutively reabsorbed and secreted, either passively or actively, by the tubular cells. The *distal convoluted tubule* is relatively impermeable but is characterized by the ability to respond to hormonal stimulation. In the presence of e.g. antidiuretic hormone (ADH), the *distal convoluted tubule* responds by increasing the reabsorption of H_2O and thereby functions to regulate the retention of H_2O during dehydration [137]. During the flow towards the bladder, the *collecting duct* channels and continuously alters the composition of the fluid by reabsorption and secretion mechanisms. Ultimately, these complex processes produce the concentrated solution of metabolic waste products known as final urine.

Although urinary elimination of proteins is limited by the nephrons, small amounts (<0.1 g/L) of urinary proteins can still be detected during normal physiological conditions [137]. This is attributed to several mechanisms, e.g. plasma proteins evading tubular reabsorption, tubular secretion of soluble proteins, detachment of GPI anchored proteins, exosome shedding or even by protein excretion from the prostate gland [141, 142]. The majority (70%) of urinary proteins originate from the kidneys and the urothelium whereas the remaining fraction (30%) is derived from the glomerular plasma filtrate [143].

Urinary Proteomics

The first proteomic study of healthy human urine dates back to 1997 where the combination of LC-MS and Edman degradation experiments identified 37 protein fragments [144]. With the improvement of proteomic and bioinformatics tools, the number of identified proteins in healthy human urine reached approximately 800 by 2005 [142, 145-147]. The following year, a single study identified more than 1500 urinary proteins in healthy individuals [148]. The largest dataset for the urinary proteome, which includes more than 2300 proteins identifications, was reported in 2009 [149]. Most recently, Mann and coworkers identified a common set of nearly 600 proteins in

healthy human urine which were defined as the “core urinary proteome”, i.e. proteins that are readily detectable in all studied individuals [150]. Notably, the 20 most abundant proteins, which account for nearly 70% of the core urinary proteome by mass, were all but one (serum albumin) found to be glycoproteins.

During the past few years, proteomic strategies have also been applied to study the urinary proteome in relation to various disease states. Urinary protein biomarkers have been successfully identified not only for renal diseases but also for other disorders in distal organs [151]. Interestingly, glycoproteins are also overrepresented in the urinary panel of biomarkers [152].

Proteomic analysis of urine remains analytically challenging mainly because of the dilute protein concentration and high content of interfering substances, i.e. salts, pigments and other metabolic waste products. Despite this fact, urine is still considered an important diagnostic medium for systemic and renal diseases. Furthermore, the complex protein composition of urine [148-150] constitutes a challenging proteome not only for biomarker discovery but also for structural characterization of human glycoproteins.

2 AIM

The general aim of this thesis was to develop glycoproteomic techniques for characterization of N- and O-linked glycopeptides. The specific aims were to:

- Develop methods for efficient enrichment of N- and O-linked glycopeptides originating from glycoproteins/peptides in complex biological samples (e.g. CSF and urine).

- Utilize established proteomic LC-MS/MS techniques and bioinformatics tools for characterization of glycoproteins by identification of peptide sequences, glycan structures and their attachment sites.

- Refine the sialic acid capture-and-release protocol to enable targeted O-glycoproteomics.

- Conduct in-depth glycoproteomic analysis of targeted glycoproteins for structural as well as (semi)quantitative analysis of different glycoforms.

3 METHODS

3.1.1 Sialic acid capture-and-release

Capture of sialylated glycoproteins

The glycopeptide enrichment strategy described in this thesis is based on two established facts: i) selective oxidation of sialic acids by periodate (IO_4^-) [153, 154] and ii) selective hydrolysis of sialic acid glycosidic bonds by mild acid treatment [155]. Periodate oxidation of *cis*-diols introduces aldehyde functionalities and, depending on the reaction conditions, vicinal hydroxyls of either sialic acids or monosaccharides in general may thus be converted to aldehydes. Aebersold and co-workers elegantly showed that serum glycoproteins, periodate oxidized under relatively harsh conditions (15 mM IO_4^-, RT, 60 min), may be coupled to a solid phase (hydrazide beads) via their aldehyde functionalities [156]. Inspired by this study, an alternative approach for targeted enrichment of sialylated glycoproteins was developed in this thesis (Fig. 7). The procedure involves selective oxidation of sialic

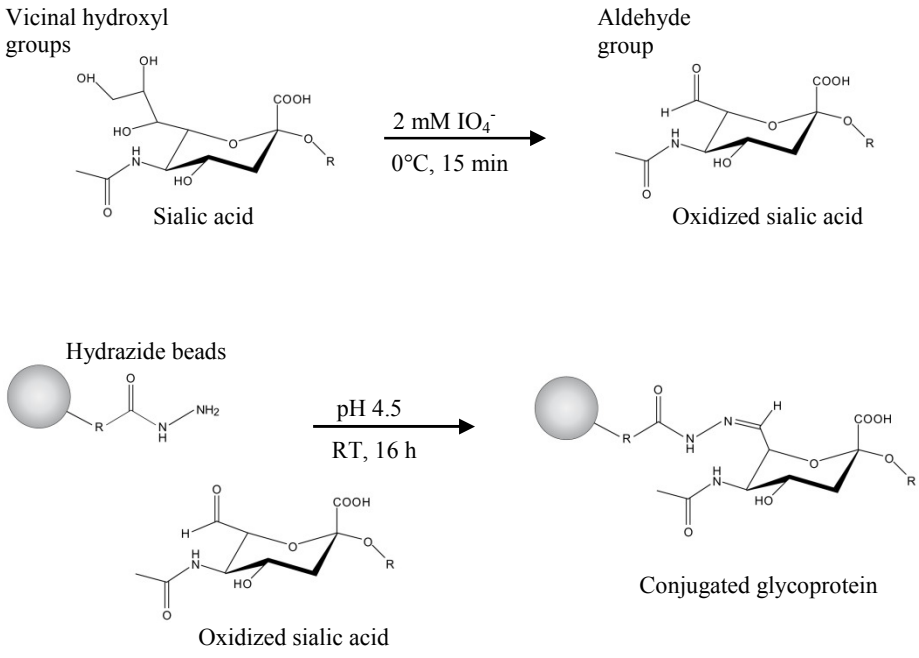


Figure 7. Capture of sialylated glycoproteins

acids by mild periodate treatment (2 mM IO_4^-, 0°C, 15 min) and coupling of oxidized sialic acids to hydrazide beads. The solid phase, conjugated with sialylated glycoproteins via stable hydrazone bonds, was subsequently washed to remove unbound proteins. Following reduction and alkylation, the conjugated glycoproteins were trypsin digested and the trypsin released peptide fraction was extracted for MS analysis. Consequently, tryptic glycopeptides, conjugated via sialylated glycans, were preserved on the hydrazide beads.

Release of desialylated glycopeptides

The solid phase was initially washed to remove residual non-glycosylated peptides. The sialic acid glycosidic bonds were hydrolyzed by mild acid treatment (0.1M formic acid, 80°C, 60 min) to release desialylated N- and O-linked glycopeptides for MS analysis. The entire enrichment procedure, described in more detail in paper **I**, **III** and **IV**, is summarized in Fig. 8.

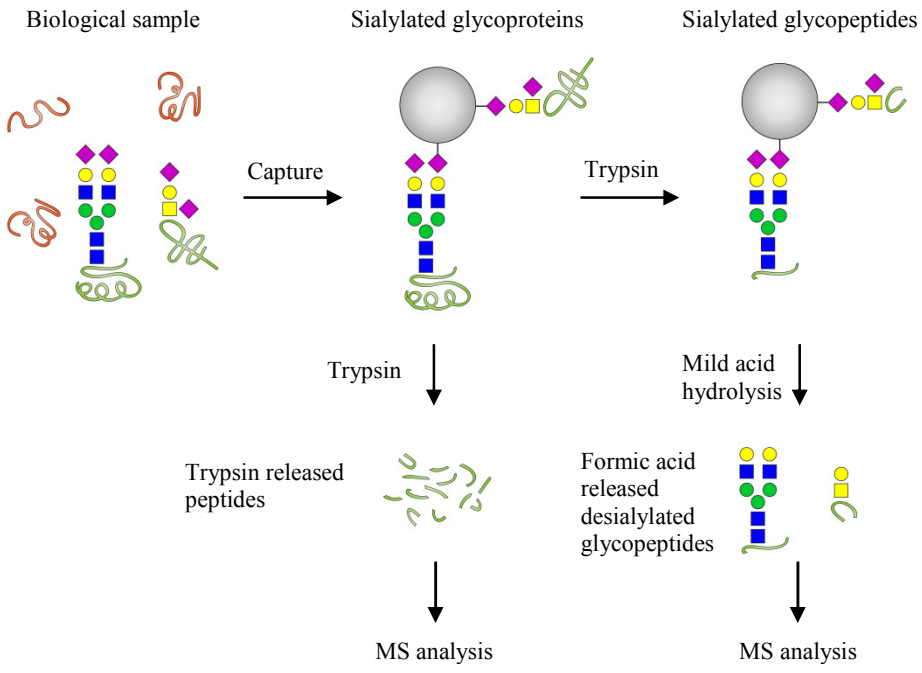


Figure 8. Sialic acid capture-and-release for enrichment of sialylated glycoproteins and desialylated N-and O-linked glycopeptides.

3.1.2 MS analysis

Glycoprotein identification

Captured glycoproteins were identified by established proteomics methods [83]. Briefly, trypsin released peptide fractions were chromatographed (reversed-phase C18 column) and analyzed on-line by ESI-LTQ-FTICR. Precursor ions were mass measured by FTICR, isolated and subjected to collision induced dissociation (CID) to induce peptide fragmentation. This fragmentation technique is performed in the LTQ and involves the collision of precursor ions with neutral helium molecules. Upon collision, the kinetic energy is partially converted to internal energy (vibrational excitation) which ultimately leads to peptide backbone fragmentation (Fig. 9) into b- and y-type ions (nomenclature according to Roepstroff and Fohlman [157]). The experimental peptide fragments were identified and grouped into protein identities by the Mascot algorithm.

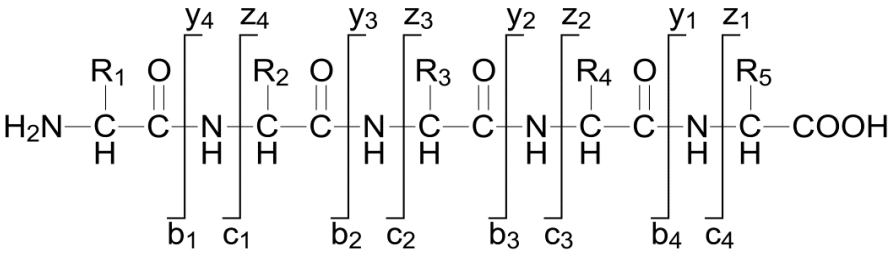


Figure 9. Nomenclature of peptide N- and C-terminal fragments.

Glycopeptide characterization by CID

The key features of glycopeptide CID fragmentation have been discussed elsewhere [158]. In summary, glycopeptide MS^2 spectra are characterized by abundant Y-type ions and diagnostic B-type oxonium ions (nomenclature according to Domon and Costello [159]) that arise from glycosidic bond fragmentations (Fig. 10). Peptide backbone fragments are usually not observed during CID MS^2 of glycopeptides.

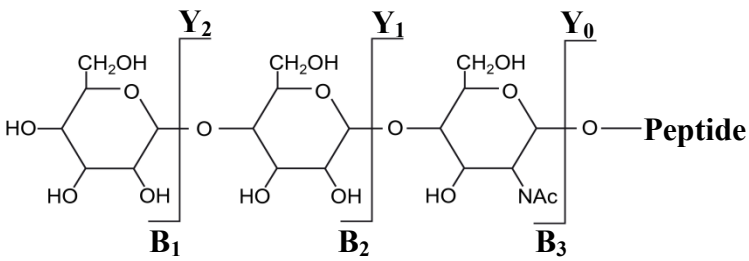


Figure 10. Nomenclature for describing the fragmentation of glycans and glycopeptides.

Here, an alternative MS strategy was employed in order to elucidate the primary structure of glycopeptides, i.e. both glycan- and peptide sequences. Glycopeptides were chromatographed and analyzed using the same MS platform as above. Precursor ions were mass measured by FTICR and fragmented by CID in the LTQ. Subsequently, the five most abundant fragment ions were isolated and subjected to additional rounds of CID-fragmentation. In other words, an MS^2 spectrum and five MS^3 spectra were acquired for individual glycopeptides. N- and O-linked glycan sequences were manually identified by tracing the loss of monosaccharides from the peptide backbone. Selected MS^3 spectra, corresponding to peptide backbone fragmentation, were identified by the Mascot algorithm.

Glycopeptide characterization by ECD

Electron capture dissociation (ECD) is an alternative technique for peptide fragmentation [160]. The principle of ECD, which is performed in the FTICR mass analyzer, is based on the reaction of multiply protonated precursor ions with low-energy electrons. In this process, electrons are captured by protonated peptides (M), resulting in the formation of peptide radicals (M•) that rapidly dissociate, before energy randomization occurs, into peptide c- and z-type fragments (Fig. 9) according to the reaction below.

$$[M + nH]^{n+} + e^- \rightarrow [M + nH\bullet]^{(n-1)+} \rightarrow \text{fragmentation}$$

One characteristic feature of ECD is the ability to fragment the peptide backbone without disrupting labile glycosidic bonds. ECD may therefore be used for glycopeptide analysis and, in particular, for identification of O-glycan attachment sites. This fragmentation technique was applied for the analysis of glycopeptides in paper **II**, **III** and **IV**. Electron transfer dissociation (ETD) is a similar fragmentation technique that is available on orbitrap instruments. This technique was applied for the analysis of O-glycopeptides in paper **IV.**

3.1.3 Targeted O-glycoproteomics

The sialic acid capture-and-release method was further refined to allow targeted enrichment of O-linked glycoproteins and O-linked glycopeptides. Prior to glycopeptide enrichment, N-linked glycans were specifically removed by enzymatic PNGase F treatment. The sample was subsequently processed according to the sialic acid capture-and release procedure described above. This strategy was initially evaluated using a well-characterized glycoprotein (bovine fetuin) and subsequently applied to a complex biological sample (CSF).

3.1.4 Glycoproteomics of APP/Aβ

APP/Aβ peptides (including APP/Aβ glycopeptides) were immunoprecipitated according to the method developed by Portelius and co-workers [131]. Briefly, the monoclonal 6E10 antibody (mouse IgG), specific for Aβ residues 4-9, was coupled to magnetic Dynabeads (Sheep anti mouse, IgG). Subsequently, 5 mL CSF was incubated with the magnetic beads to allow immunoprecipitation of Aβ peptides. The APP/Aβ peptides were eluted by 0.5% formic acid and evaporated to dryness prior to MS analysis.

APP/Aβ peptides were dissolved and chromatographed on a reversed-phase C4 column. Precursor ions were mass measured by FTICR and characterized by CID and ECD fragmentation techniques. The MS data was annotated manually.

4 RESULTS

4.1.1 Paper I

The sialic acid capture-and-release method for enrichment of glycopeptides was initially evaluated using two model glycoproteins, i.e. human transferrin (TF) and bovine fetuin (FET). The efficient enrichment of tryptic N-glycosylated peptides (N-glycopeptides) from TF is shown in Fig. 11A. The base peak ion chromatogram revealed three dominating peaks, eluting at 30-40 min, which were mass measured by FTICR (Fig. 11B). The molecular ion masses were matched, with good mass accuracy (±10 ppm), to tryptic TF peptides with five Hex and four N-acetylhexosamine (HexNAc) units, tentatively assigned as desialylated biantennary complex type N-glycans. The peak at ~36 min (m/z 1028.5) appeared due to the deamination of the N-terminal Cys residue.

The glycan composition of each precursor ion was characterized by CID fragmentation (Fig. 11C). In CID-MS^2, diagnostic B-type oxonium ions (m/z 366 and m/z 528) and abundant Y-type fragments revealed structural information on the N-glycan sequence. The presence of relatively abundant Y_1-type ions (m/z 840.9), corresponding to peptide+HexNAc fragments, was found to be a common feature during CID-MS^n of N-glycopeptides in

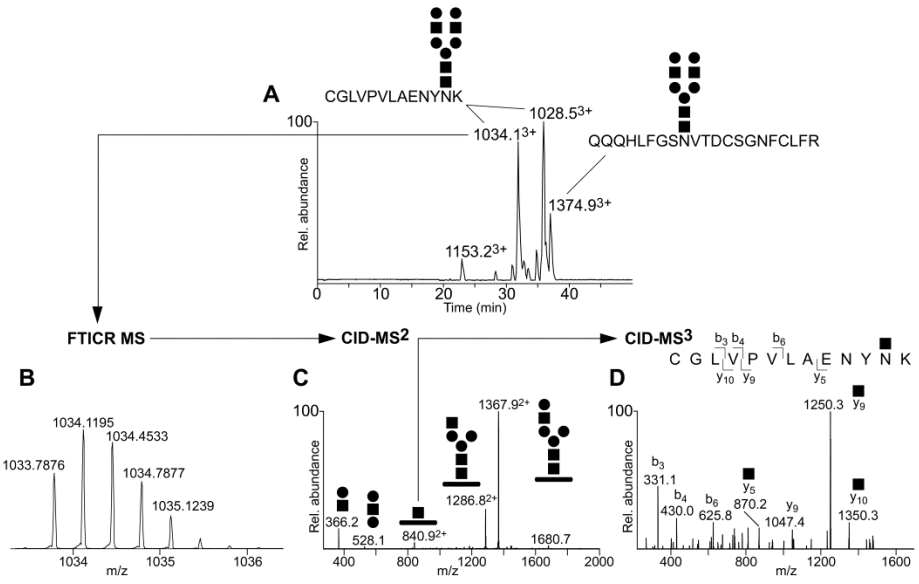


Figure 11. MS analysis of sialic acid capture-and-release enriched glycopeptides. (A) Base peak ion chromatogram of transferrin N-glycopeptides. (B)FTICR MS of the molecular ion eluting at ~32 min. (C) CID-MS^2 for N-glycan sequence analysis. (D) CID-MS^3 for peptide sequence analysis.

accordance with previous studies [100]. CID-MS^3 fragmentation of Y_1-type ions generated peptide backbone b- and y-type fragments (Fig. 11D) which were used to identify the peptide sequence ($p<0.05$ significance level) through the Mascot algorithm. Bovine fetuin was enriched and analyzed according to the same principle as above. All three N-linked glycopeptides of FET were characterized. In addition, three glycopeptides with HexHexNAc core 1-like O-glycans were identified (Supplemental data, paper **I**). Taken together, these experiments demonstrated that both N- and O-glycopeptides may be efficiently enriched by the sialic acid capture-and-release method. Also, the CID-MS^n approach was found to be a successful strategy for identifying both glycan and peptide sequences of N- and O-glycopeptides.

Next, the sialic acid capture-and-release procedure was applied to a complex biological sample, i.e. CSF. Bottom-up proteomic analysis of the trypsin released peptide fraction (Fig. 8) identified 84 proteins, the majority of which were found to be glycoproteins. The base peak ion chromatogram of the formic acid released glycopeptide fraction was dominated by various N- and O-linked glycopeptides (Fig. 12A), showing that the sialic acid capture-and-release method was applicable to complex biological samples.

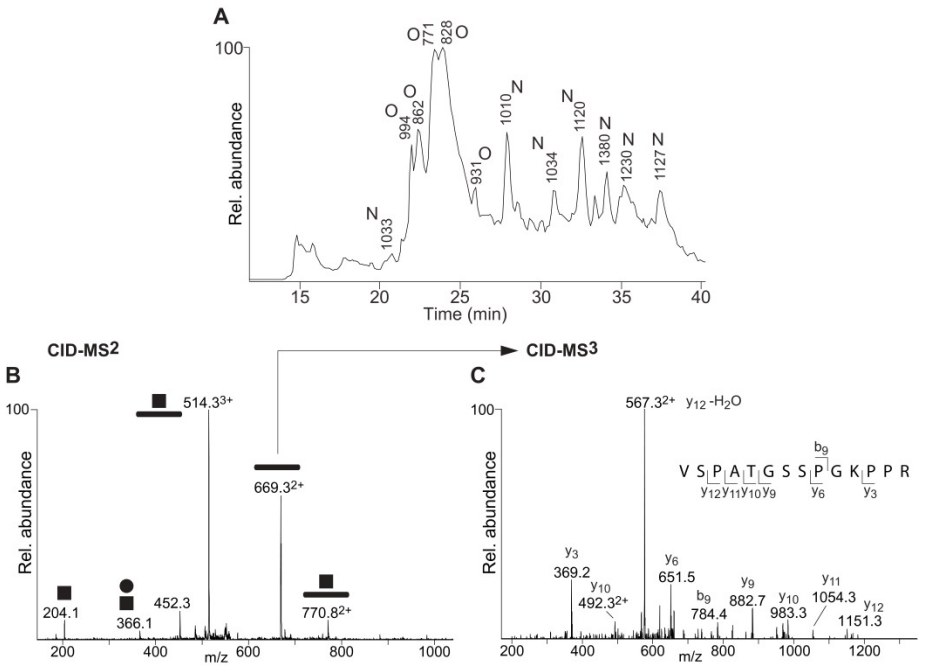


Figure 12. (A) Base peak ion chromatogram of sialic acid capture-and-release enriched glycopeptides from CSF. (B) CID-MS^2 analysis for sequencing O-glycans. (C) CID-MS^3 spectrum of the Y_0 fragment from panel B used for identification of the peptide sequence (cystatin C).

In total, 36 N- and 43 O-glycopeptides were characterized with respect to glycan and peptide sequences. The $Hex_5HexNAc_4$ composition, tentatively assigned as the biantennary complex type N-glycan structure, was found to be the most common glycoform of N-glycopeptides in CSF. The enriched O-glycopeptides were only detected with core 1-like HexHexNAc O-glycans. The majority of these (37 unique peptides) were identified as new, previously unknown O-glycopeptides, e.g. the peptide from cystatin C (Fig. 12B-C).

4.1.2 Paper II and IV

In the follow-up study to paper **I**, the enrichment method was refined to allow targeted O-glycoproteomics of complex biological samples. Initially, PNGase F treated FET was enriched and analyzed according to the procedures described above. O-glycopeptides, but not N-glycopeptides, were identified in the following MS analysis (not shown). Subsequently, CSF was PNGase F treated and enriched by the sialic acid capture-and-release method. As expected, the base peak ion chromatogram (Fig. 13A) lacked the characteristic N-glycopeptide peaks observed in paper **I** and was instead dominated by a prominent peak at ~23 min (m/z 994, Apolipoprotein E O-glycopeptide). The absence of CSF N-glycopeptides facilitated characterization of many low-abundant O-glycopeptides, e.g. the precursor ion (m/z 838) eluting at ~22 min (Fig. 13B). This precursor ion was identified as the APP peptide GLTTRPGSGLTNIK with 3 HexHexNAc core 1-like O-glycans. The peptide sequence was found to be located approximately ~20 residues N-terminally to the β-secretase cleavage site of APP.

CID-MS^n analysis revealed the majority of O-glycopeptides to be substituted with core 1-like HexHexNAc O-glycans. ECD analysis allowed the precise identification of O-glycan attachment sites (Fig. 13C). In total, 78 O-glycopeptides, constituting 95 putative O-glycosylation sites, were characterized with respect to glycan and peptide sequences. The combination of CID and ECD fragmentation techniques pinpointed 67 HexHexNAc core 1-like O-glycans to unique Ser/Thr attachment sites. This approach, discussed in more detail in paper **IV**, demonstrated that scarce O-glycopeptides are made accessible for glycoproteomic characterization by PNGase F treatment and sialic acid capture-and-release enrichment. In addition, a very interesting APP O-glycopeptide was identified which justified a more detailed, in-depth O-glycoproteomic analysis of this glycoprotein.

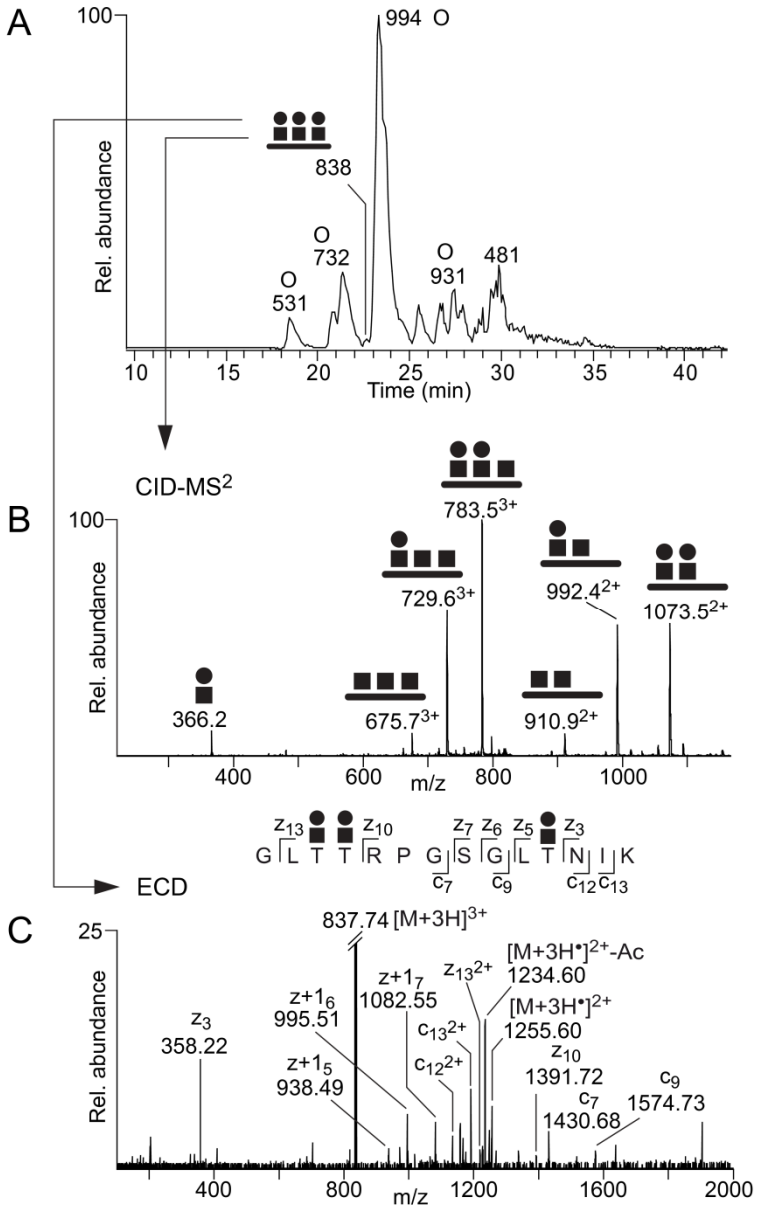


Figure 13. Targeted O-glycoproteomics of CSF. (A) Base peak ion chromatogram of sialic acid capture-and-release enriched O-glycopeptides. (B) CID-MS^2 analysis of the APP O-glycopeptide at m/z 838. (C) ECD fragmentation of the precursor ion at m/z 838 for identification of O-glycan attachment sites.

To achieve this, APP/Aβ peptides were immunoprecipitated from human CSF samples with the 6E10 antibody (paper **II**). Previously, this approach was used to identify various Aβ1-X and APP/AβX-15 peptides in CSF [131, 132]. These results were successfully reproduced; 33 Aβ1-X peptides and 14 APP/AβX-15 peptides were identified in the 6E10 immunoprecipitated CSF fractions by MS analysis. Co-eluting with specific APP/AβX-15 peptides, several molecular ions with mass increments corresponding to $(Sia)_2HexHexNAc$ and/or $(Sia)_{1-4}(HexHexNAc)_{1-4}$ glycans were also observed. CID-MS^2 analysis of these precursor ions corroborated the presence of sialylated core 1-like O-glycans on APP/AβX-15 peptides (Fig. 4, paper **II**). The O-glycans were mapped, by CID and ECD analysis, to specific Thr residues located at positions -39, -21, -20 and -11 relative to Asp1 of the Aβ sequence. The GLTTRPGSGLTNIK sequence (Fig. 13), which constitutes the -23 to -10 region of APP/AβX-15 peptides, was thus confirmed to accommodate three sialylated core 1-like O-glycans at Thr(-21, -20 and -11). The attachment site for one of the APP/AβX-15 O-glycans, located at either Thr(-9) or Ser(-5), could not be determined. In all, 37 unique APP/AβX-15 O-glycopeptides were identified (Table S1, paper **II**).

Next, the molecular ions co-eluting with the Aβ1-X peptides were investigated in more detail. Several precursor ions, with m/z values matching specific Aβ1-X peptides with $(Sia)_{1-3}HexHexNAc$ glycans, were observed in the FTICR mass spectra. Prominent Y-type fragments and characteristic B-type oxonium ions verified the mono- (SA), di- (SA_2) and trisialylated (SA_3) HexHexNAc glycan sequences in the following CID-MS^n experiments. The SA_3 glycoform (Fig. 14A) was found with the unusual disialosyl sequence, which occasionally also was observed in the lactonized form (Fig. 1, paper **II**). These results indicated that the terminal Hex-linked structure was a Neu5Acα2-8Neu5Ac epitope. Additional structural heterogeneity associated with the SA_3 glycoform included O-acetylation of the disialosyl epitope. In total, 27 unique Aβ1-X O-glycopeptides were identified (Table S1, paper **II**).

The Aβ1-38 and Aβ1-40 peptide isoforms were the most abundant Aβ components of 6E10 immunoprecipitated CSF samples. However, these peptide isoforms were not observed as glycosylated. In fact, only C-terminally truncated Aβ peptides, the longest being Aβ1-20, were found to be glycosylated.

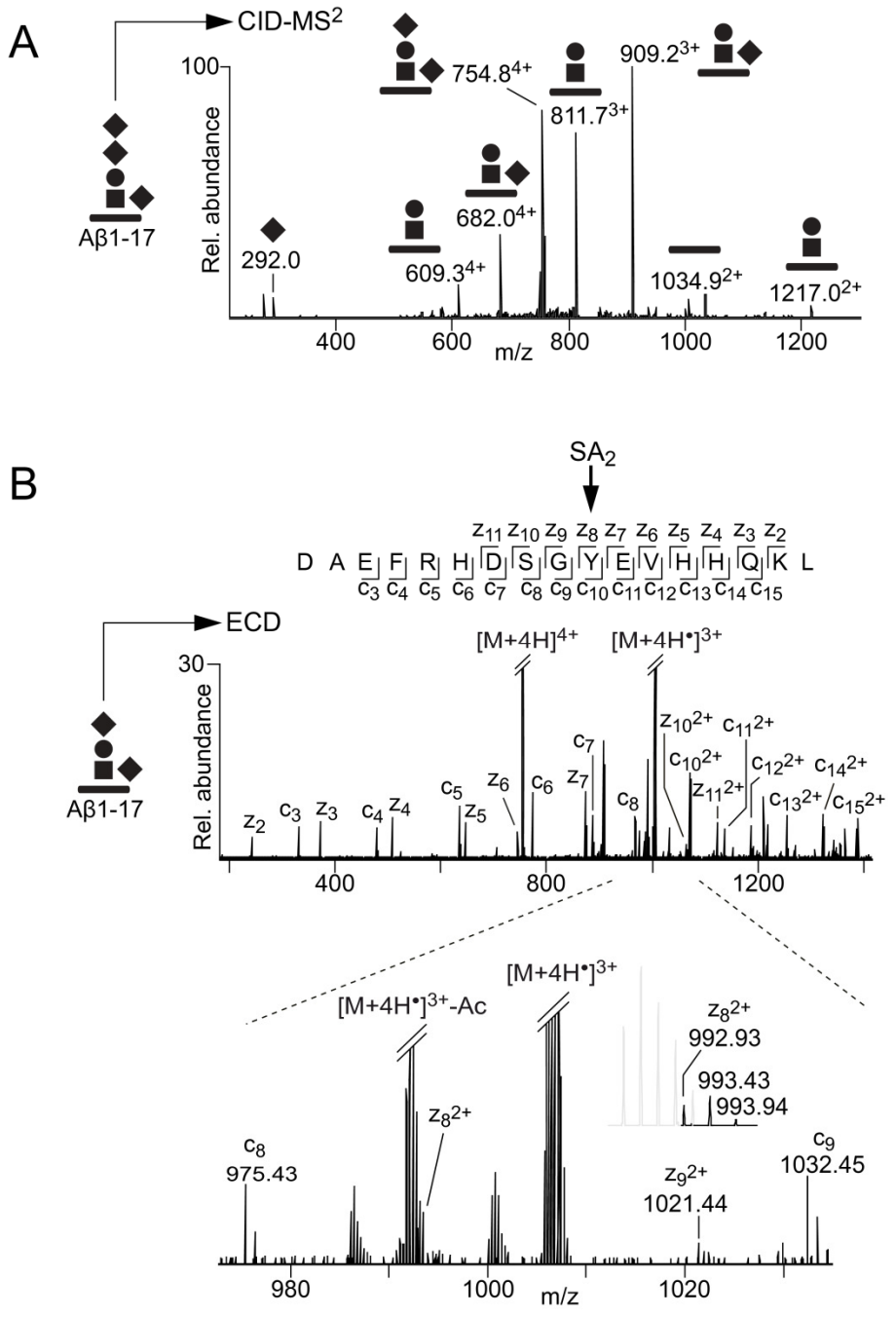


Figure 14. Tandem MS analysis of Aβ1-17 glycopeptides. (A) CID-MS^2 spectrum of the trisialylated glycoform. (B) ECD spectrum of the disialylated glycoform. The m/z expansion shows the critical c- and z-ions that pinpoint the glycosylation site to Tyr10. The isotopic peaks of the z8 fragment are resolved in the Inset.

In the following experiments, ECD was performed to verify the peptide sequences and the O-glycan attachment site of Aβ1-X glycopeptides. Given that Aβ1-20 (DAEFRHDSGYEVHHQKLVFF) contained only one possible glycan attachment site, it was naturally assumed that all Aβ1-X glycopeptides were glycosylated on Ser8. However, the ECD analysis clearly showed that the O-glycans were selectively attached to Tyr10 (Fig 14B). ECD fragmentation of the Aβ1-17 SA_2 glycoform generated a continuous c3-c15 and an uninterrupted z2-z11 ion series. The c8 (DAEFRHDS, m/z 975.43) and c9 (DAEFRHDSG, m/z 1032.45) fragment ions, which include Ser8 but exclude Tyr10, were observed without the mass of the SA_2 O-glycan whereas e.g. c10 and c11 fragment ions were found to be glycosylated. From the opposite direction, the z8 (YEVHHQKL, m/z 992.94) and z9 (GYEVHHQKL, m/z 1021.44) fragment ions, which include Tyr10 but exclude Ser8, were both observed with the mass of a SA_2 O-glycan, clearly showing that Aβ1-17 is glycosylated at Tyr10. The glycans of other Aβ1-X glycopeptides, e.g. Aβ1-15, were also mapped to Tyr10 by ECD experiments (Fig. 2 and Fig. S4, paper **II**). Taken together, these results demonstrated that the glycan attachment site of Aβ1-X glycopeptides was Tyr10. Finally, the relative signal intensities of individual Aβ1-X peptides and Aβ1-X glycopeptides, normalized to the summed intensities of all Aβ peptides (except Aβ-42), was compared for AD (n=6) and non-AD patients (n=7). A relative increase, by a factor of 1.1 to 2.5, of glycosylated versus non-glycosylated Aβ peptides was observed for the AD patients (Fig. 3, paper **II**).

In summary, the glycoproteomic analysis of 6E10 immunoprecipitated CSF samples identified five unique sialylated core 1-like O-glycans in the vicinity of the β-secretase cleavage site of APP. In addition, a unique, trisialylated HexHexNAc O-glycan, was found to reside between the α- and β-secretase cleavage sites. Tyrosine, which represents a novel attachment site for sialylated O-glycans in mammalian proteins, was found to be the modified residue of Aβ1-X glycopeptides.

4.1.3 Paper III

In this study, the sialic acid capture-and-release method was applied to an alternative, but also more challenging, biological fluid, i.e. urine. Initially, efficient enrichment of urinary glycoproteins was hampered by the high content of reducing waste products interfering with the periodate oxidation. However, these interfering compounds were successfully removed by repeated dialysis of which one was in 1.5% SDS. Following glycopeptide enrichment, 58 N- and 63 O-glycopeptides, constituting 57 putative O-glycosylation sites, were characterized by CID and/or ECD analyses. The

precise attachment site of 40 Ser/Thr linked core 1-like O-glycans was defined, of which 29 had not been previously reported. Tyrosine linked glycans were not observed for the urinary O-glycopeptides. In addition to $Hex_5HexNAc_4$-N-Asn modified glycopeptides, several other N-glycoforms were characterized, including branched and fucosylated sequences ($dHex_1Hex_7HexNAc_6$-N-Asn), tentatively assigned as tetra-antennary N-glycans. Structural O-glycan variation for specific O-glycopeptides was also observed, e.g. fucosylation of core 2-like sequences (Fig. 15) and secondary modifications (sulfation) of O-glycans. This study demonstrated that sialic acid capture-and-release enrichment, in combination with efficient protein purification techniques, may be used for glycoproteomic characterization of analytically challenging biologic samples.

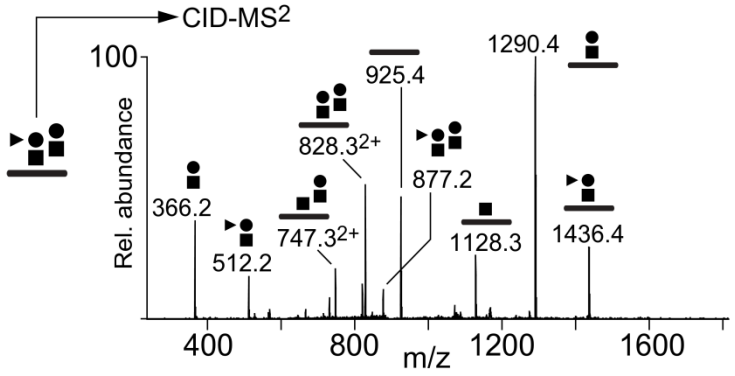


Figure 15. CID-MS^2 analysis of the urinary AVAVTLQSH O-glycopeptide (protein YIPF3) substituted with a fucosylated core 2-like glycan.

5 DISCUSSION

Glycoproteins, in contrast to DNA and nascent proteins, are secondary gene products synthesized by non-template driven processes. Thus, there are no biological blueprints for predicting the structural details of glycoproteins. Traditionally, researchers have relied on various analytical techniques, the primary being MS-based analysis, for structural characterization of glycoproteins. However, the analytical difficulties associated with the structural complexity of glycoproteins have obligated researches to adopt a certain degree of reductionism in their MS-based analyses [161].

The standard analytical approach has been to separate glycans from proteins and analyzed these components individually by either glycomic or proteomic techniques. As with any other analytical approach, this type of reductionism has its own advantages and disadvantages. In glycomics, sensitive characterization of glycan structures is achieved but the site-specific location of these glycans and the identities of their protein carriers are lost. Similarly, hundreds of glycoproteins may be identified by proteomic analysis but this approach is unable to provide any information on the actual glycosylation status of the identified glycoproteins. In other words, the classical trade-off in MS-based analysis has been glycan structures versus glycoprotein identities.

So far, MS-based strategies aimed at analyzing intact glycopeptides have generally been limited to the monoproteic approach, a definition recently used by Dodds *et al.* to describe the glycoproteomic analysis of single glycoproteins [161]. Using this approach, glycan structures and their attachment sites may be determined but this type of analysis is only applicable to simple glycoprotein mixtures.

The lack of efficient enrichment procedures has been one of the primary obstacles in unbiased glycoproteomic analyses. Accordingly, one of the primary ambitions in this thesis was to develop methods capable of efficient enrichment of N- and O-glycopeptides for glycoproteomic characterization of complex biological samples.

Sialic acid capture-and-release

The glycopeptide enrichment strategy described in this thesis capitalizes on the characteristic chemistry of sialic acids. This strategy was chosen for one important reason; the oligosaccharides of secreted and membrane bound glycoproteins are frequently terminated with sialic acids (Fig 2). By specifically targeting the sialylated glycoproteome both N- and O-

glycoproteins may thus be enriched. In contrast to lectin-based methods, which are limited by the binding specificity of the lectin(s), e.g. jacalin recognizing Galβ1-3GalNAcα-O- structures [107], this strategy is able to enrich N- and O-glycoproteins without being restricted to a particular sub-class of sialylated glycans. However, the obvious disadvantage is that non-sialylated structures, e.g. blood group antigen terminated O-glycans (Fig. 2) or high-mannose type N-glycans, are not enriched by this method.

Given that aldehyde groups are rarely found on proteins or glycoproteins, mild periodate oxidation of sialic acids (Fig. 7) is a simple way of introducing reactive aldehyde functionalities which may be utilized for targeted enrichment of sialylated glycoproteins. Under these mild conditions, the periodate oxidation is specific [154], i.e. oxidation takes place primarily at the glycerol side chain (C7-C9) of sialic acids. By comparison to the procedure of Aebersold and co-workers [156], the mild periodate treatment is also non-destructive for the underlying glycan, i.e. the ring structure of the monosaccharides is preserved. However, biological modifications at the C7-C9 hydroxyls of sialic acids, e.g. O-acetylation, preclude the mild periodate oxidation. In other words, an unmodified glycerol side chain is a prerequisite for the introduction of aldehyde functionalities. Thus, glycoproteins modified by e.g. terminal OAcNeu5Ac structures are inaccessible for enrichment by this procedure.

The aldehyde groups of oxidized sialic acids react with hydrazide beads to form covalent hydrazide bonds that conjugate sialylated glycoproteins to the solid phase (Figs. 7 and 8). The covalent linkage confers a significant advantage over traditional enrichment strategies, e.g. HILIC or lectin-based methods, by allowing relatively harsh washing conditions (e.g. >1.5M NaCl, >1% SDS or >50% acetonitrile) to be employed. Such conditions are inapplicable to conventional enrichment strategies based on non-covalent interactions. Considerable reduction of sample complexity, i.e. removal of non-glycosylated proteins, with concomitant purification of sialylated glycoproteins is thus achieved by repeated washing of the solid phase.

The following trypsin digestion of the conjugated glycoproteins fulfills two functions: i) it releases non-glycosylated peptides from the solid phase (which may be identified by proteomic methods) and ii) it generates tryptic glycopeptides on the solid phase. The enrichment method is thus compatible with proteomic strategies aimed at identifying sialylated glycoproteins in complex biological samples. In paper **I** and **III**, the trypsin released peptide fraction was analyzed by bottom-up proteomics (section 1.2.2) which

confirmed that glycoproteins are efficiently enriched by the sialic acid capture-and-release method (table S1, paper **I** and table S1, paper **III**).

Trypsin digestion of the conjugated glycoproteins leaves behind covalently linked glycopeptides, which is beneficial for several reasons. Perhaps the most important reason is however that this procedure allows conjugated glycopeptides to be separated from non-glycosylated peptides by sequential washing of the solid phase. Thus, ion suppression of glycopeptides by co-eluting non-glycosylated peptides, which may have a dramatic impact on the analytical sensitivity [162], is avoided in downstream MS analyses. Ion suppression of glycopeptides is one of the primary limitations in glycoproteomic analysis. In monoproteic glycoproteomics, the tryptic digest (peptides and glycopeptides) of purified glycoproteins is usually analyzed with or without additional enrichment at the glycopeptide-level. Although this strategy is successful for purified glycoproteins [163], the ion suppression effect exerted by non-glycosylated peptides becomes a significant issue for complex protein mixtures. A plausible explanation for the lack of unbiased glycoproteomic studies on biological samples might thus be attributed to the inability of conventional enrichment methods, based on non-covalent interactions, to efficiently separate non-glycosylated peptides from glycopeptides.

The conjugated glycopeptides are subsequently released by mild acid treatment (Fig. 8). The release mechanism is specific for sialic acids [155], i.e. glycosidic linkages between sialic acids and penultimate monosaccharides, but not between dHex, Hex or HexNAc residues, are hydrolyzed. In other words, glycopeptides are released from the solid phase as desialylated species. Clearly, the loss of sialic acids during the release procedure is one of the primary limitations of this method. No conclusions may thus be drawn with respect to sialic acid heterogeneity, except for the fact that the glycopeptides were sialylated to some degree. However, the main advantage is the ability to isolate intact glycopeptides. By comparison, Aebersold and co-coworkers used PNGase F to release of formerly N-glycosylated peptides [156]. Although the enzymatic release is specific and allows identification of the N-glycan attachment site (based on the conversion of Asn to Asp at the glycosylation site), this procedure removes the entire N-glycan and is only applicable to N-glycopeptides. In contrast, the mild acid treatment preserves the glycan structure and works equally well for both N- and O-glycopeptides.

In short, the specificity of the enrichment procedure described in this thesis is defined by two steps: i) selective oxidation of sialic acids by mild periodate

treatment and ii) selective hydrolysis of sialic acid glycosidic bonds by mild acid treatment. Taken together, these steps enable selective enrichment of desialylated glycopeptides for MS analyses. Consistent with this statement, glycopeptides were indeed found to be the dominating components in the formic acid released fractions of transferrin (Fig. 11A), CSF (Figs. 12A and 13A) and urine (Fig. S4, paper **III**) samples. Considering that transferrin alone is trypsinized into ~80 protein fragments, the appearance of three dominating peaks in the base peak chromatogram demonstrated the efficient enrichment of TF N-glycopeptides. Notably, even contaminants (β-2-glycoprotein N-glycopeptides, m/z 1153.2, ~23 min, Fig. 11A) in the ≥98% pure TF sample were enriched.

It should be stressed that a similar enrichment strategy, termed reverse-glycoblotting, was independently developed by Nishimura and co-workers. In their initial approach, oxidized sialic acids of tryptic glycopeptides were conjugated to aminooxy-functionalized polymers via stable oxime bonds and released by trifluoroacetic acid treatment (hydrolysis of sialic acid glycosidic bonds) [164]. Although the chemistry is similar to the sialic acid capture-and-release method, their initial study on mouse serum was relatively unsuccessful (only 6 N-glycopeptides identified). In 2010, an improved enrichment strategy, based on hydrazide chemistry, was reported by the same group [108]. In their refined method, release of sialylated glycopeptides by ice-cold 1M HCl was reported, along with the identification of 67 N-glycopeptides from mouse serum. Thus, the unique release strategy of Nishimura and co-workers represents an alternative method for enrichment of sialylated glycopeptides. It is reasonable to believe that this strategy is also suitable for enrichment of sialylated O-glycopeptides but such data has not been presented by these authors.

MS analyses

The MS analyses described in this thesis were based on accurate FTICR mass measurements followed by tandem MS experiments utilizing primarily CID and ECD fragmentation techniques. Although these methods are powerful tools for qualitative studies on glycopeptides, MS analyses alone are unable to resolve the absolute structure of glycopeptides, at least not according to the procedures described in this thesis. Regardless of mass accuracy and resolution, a mass measurement of two stereoisomers (section 1.1.1) will always yield the same mass, e.g. the oxonium ions of GalNAc and GlcNAc at m/z 204. Similarly, the anomeric configuration, linkage position and monosaccharide identities of disaccharides will not be resolved by simple mass measurements. For example, the oxonium ions of LacNAc (Fig. 1) and core 1 (table 1) structures will both appear at m/z 366, even though these

disaccharides are structurally unique. This limitation, which is valid for both B- and Y-type fragments (Fig. 10), is recognized by the use of non-colored symbols for glycan annotations in this thesis.

The CID-MS^n approach was utilized as the primary method for the identification of glycopeptides in unbiased glycoproteomic analyses. This MS strategy is based on the fact that glycans may be sequence in CID-MS^2, where predominant fragmentation of glycosidic bonds is expected [158], followed by CID-MS^3 of Y_1- or Y_0-type ions, where peptide b- and y-type fragmentations are induced. In other words, the CID-MS^2 event is used to "deglycosylate" the peptide component whereas the CID-MS^3 event is employed to fragment the peptide component itself. Selected CID-MS^3 spectra, with peptide specific b- and y-type fragments, may therefore be queried for protein identification by the Mascot algorithm. In this way, glycopeptides may be characterized and confidently identified by utilizing established proteomic platforms. However, since individual precursor ions are subjected to multiple fragmentations events, the duration of each CID-MS^n experiment is relatively time-consuming by comparison to standard proteomic methods. Thus, the primary disadvantage of the CID-MS^n approach is that fewer precursor ions are analyzed during each chromatographic time-frame.

During the course of data analysis, we concluded that the Mascot algorithm was unable to identify specific O-glycopeptides enriched from CSF (paper **I**). By relaxing the search parameters to allow semitryptic O-glycopeptides to be considered in the database search, the Mascot algorithm successfully identified several peptides with C- or N-terminal non-tryptic ends, e.g. the R.GVEYVCCPPPGTPD.P peptide from Amyloid-like protein 1. The non-tryptic C-terminus is most likely the result of the formic acid treatment, which is known to hydrolyze peptides at Asp residues [165]. Similarly, further relaxing the Mascot search by querying the NCBInr database, a redundant peptide sequence database which includes single-nucleotide polymorphisms (SNPs), allowed for the identification of the cystatin C O-glycopeptide shown in Fig. 12. It was thereby concluded that relaxed database searches are beneficial in unbiased glycoproteomic analyses. Notably, the A.VSPATGSSPGKPPR.L semitryptic sequence (A→T mutation) was found to originate from the N-terminal part of cystatin C. This peptide stretch includes the VSPATG sequence, which normally is part of the signal peptide that is enzymatically removed during the maturation of cystatin C [166]. It is tempting to speculate that the A→T mutation allows a sialylated O-glycan to be introduced at the Thr residue which, in turn, alters the proteolytic processing of this protein. However, whether or not this is the

case remains to be investigated. Nevertheless, these results tend to suggest that sialylated O-glycans are able to suppress proteolytic processing of proteins, in agreement with previous studies [43].

Targeted O-glycoproteomics

These interesting results inspired us to refine our enrichment strategy for targeted O-glycoproteomic analysis. As described above, this was achieved by including the PNGase F treatment prior to the sialic acid capture-and-release enrichment of CSF (paper **IV**). The enzymatic release of N-glycans from CSF glycoproteins allowed for specific enrichment of O-glycopeptides (Fig. 13A). However, the most abundant O-glycopeptides (hemopexin, m/z 862, m/z 771 and m/z 828, Fig. 12A) characterized in paper **I** were not recovered after the PNGase F treatment. Presumably, the removal of N-glycans affects the solubility of hemopexin and causes it to precipitate before the oxidized sialic acids of the O-glycopeptides react with the hydrazide beads. This limitation may be circumvented by performing the PNGase F digestion after the capture of sialylated glycoproteins (not shown).

The removal of N-glycans, and thereby the analytical interference of N-glycopeptides, resulted in the identification 95 putative O-glycosylation sites, of which 67 were mapped to unique Ser/Thr residues by combined CID, ECD and ETD analyses. By surveying the amino acid sequence surrounding the O-glycan attachment sites, we noted that proline residues are prominent in the -1, +1 and +3 positions relative to the glycosylated Ser/Thr residues. These results are in agreement with previous studies [167, 168] and may be used to improve current bioinformatics tool for predicting O-glycan attachment sites. Given that the majority of these O-glycopeptides were found to be modified with core 1-like HexHexNAc-O-Ser/Thr sequences, efforts were also made to automate the identification procedure for these O-glycopeptides (paper **IV**). Although still under development, this type of data analysis holds great potential for rapid identification of HexHexNAc glycosylated peptides, i.e. what might take weeks of manual annotation is automatically identified in less than five minutes by this procedure. However, the identification of peptide glycoforms exceeding the simple core 1-like is not easily automated. This area of research requires the combined efforts of bioinformaticians and glycobiologists if further progress is to be made.

Recently, Clausen and co-workers presented a unique strategy for targeted O-glycoproteomics of cell lines [109]. In their approach, zinc finger nuclease (ZFN) gene targeting of the molecular chaperone Cosmc (Fig. 3) was used to generate cells expressing the truncated Tn- and sialyl-Tn antigens. Following lectin affinity enrichment, more than 350 O-glycosylation sites were

identified by MS analyses. Although this method is limited to cell-based studies, the ZFN approach is presently the most efficient method for profiling O-glycan attachment sites in complex protein mixtures. Notably, only twelve of the O-glycan attachment sites reported in paper **IV**, including the three Thr O-glycosites in Fig. 13C, were found to be in common with this study. For the O-glycopeptides identified in paper **III**, only two O-glycosites were found to be in common with the study of Clausen and co-workers. This surprisingly low overlap tends to suggest that the O-glycoproteome might be larger than previously believed.

Considering that APP is implicated in the disease progression of AD (section 1.3.1), the APP O-glycopeptide identified by the targeted O-glycoproteomics approach (Fig. 13) was perhaps the most interesting finding in paper **IV**. The identified HexHexNAc-O-Thr glycans, located in proximity to the β-secretase cleavage site of APP, provoked many speculations, including the possibility that additional O-glycans might be located in the immediate vicinity or even within the Aβ sequence itself. These questions were addressed by the glycoproteomic analysis of immunoprecipitated APP/Aβ peptides from CSF (paper **II**).

Glycoproteomics of APP/Aβ

Immunoprecipitation of APP/Aβ peptides was employed as an alternative approach to study the O-glycosylations of APP. One of the primary advantages of this method, by comparison to the sialic acid capture-and-release enrichment, is the ability to analyze sialylated O-glycopeptides. Accordingly, 64 unique O-glycopeptides of variable peptide lengths, sialic acid heterogeneities, sialic acid modifications and site occupancies were characterized. The relative amounts of specific APP/AβX-15 peptides and glycopeptides were found to vary with respect to their glycosylation status, indicating that O-glycosylations may suppress the proteolytic processing of APP at e.g. the -25 region (Fig. S9, paper **II**). Potentially, this argument may also apply to the β-secretase enzyme (section 1.3.1).

In addition to the O-glycosylations of APP/AβX-15 peptides, a unique trisialylated O-glycans was also found to reside within the Aβ sequence itself, but only on shorter Aβ1-X peptides, with X being equal to or less than 20. This tends to suggest that APP processing, in the presence of Aβ glycosylation, is altered, resulting in a shift towards the non-amyloidogenic pathway (section 1.3.1). This conclusion is supported by the lack of glycosylation on longer Aβ peptides, e.g. Aβ1-42. Most likely, the presence of a sialylated O-glycans within the Aβ sequence is able to influence the way

by which this region of APP is recognized and processed by α-, β- and γ-secretases.

The CID and ECD experiments also revealed that Aβ1-X peptides were glycosylated at Tyr10, in contrast to the anticipated Ser8. Given that sialylated O-glycans have previously only been described for Ser/Thr residues (section 1.1.2), the ECD analysis in Fig. 14B thus identified Tyr as a novel attachment site for sialylated O-glycans in human proteins.

The most abundant Aβ1-X glycopeptides were calculated to be 10-30 pg/mL (100-200 pg/mL for the non-glycosylated variants), based on the signal intensity of an isotopic internal standard. In the context of analytical sensitivity, considerable depth, at the expense of breadth, was thus achieved. In the pilot study of AD and non-AD patients, a relative increase of Tyr10 glycosylated versus non-glycosylated Aβ peptides was observed for the AD patients. Thus, the Tyr10 glycosylated Aβ peptides might be used as early indicators (biomarkers) for the progression of AD. However, the small patient groups are obviously the primary limitation of this study, and these results should be confirmed in larger prospective studies to establish Aβ glycopeptides as clinically useful biomarkers. Notably, Tyr10 glycosylated Aβ peptides were recently also identified in cat CSF (Brinkmalm *et al*., submitted manuscript). These findings indicate that the cellular machinery responsible for synthesizing complex tyrosine glycosylation may be present in all species belonging to the mammalian magnorder Boreoeutheria. Animal models may therefore be used not only for general studies on complex tyrosine glycosylations but also for AD-related studies on Aβ glycosylations.

Aβ peptides are constitutively produced and eliminated during non-pathological conditions (section 1.3.1) but it is still unknown why Aβ peptides, under certain conditions, precipitate into amyloid plaques in humans. Studies in murine models show that overexpression of human APP results in amyloid deposition whereas wild-type mice do not develop amyloid plaques with age [169]. The amino acid sequences differ only by 4% for human and murine APP and three of these residues are located within the Aβ sequence itself (R5→G, Y10→F and H13→R). Recently, Kummer *et al*. demonstrated that Tyr10 nitration of Aβ critically enhances the amyloidogenic properties of Aβ1-42 [170]. These authors argued that Tyr10 nitration, resulting from inflammatory processes in the brain, might be responsible for some of the early events that trigger amyloid deposition. This may explain why human Aβ with Tyr10, and not murine Aβ with Phe10, precipitates into amyloid plaques in animal models. In keeping with this, it is interesting to note that the reactive properties of tyrosine residues are, to a

great extent, governed by the electron donating properties of the phenolic hydroxyl group. By comparison to the benzene ring of phenylalanine, the phenolic hydroxyl group renders tyrosine 1000 times more reactive in aromatic nitration [171]. The electron donating properties, and thereby the reactivity of the tyrosine residues, is also affected by substitutions at the phenolic hydroxyl group. It is therefore reasonable to believe that nitration of Aβ Tyr10 residues may, according to the same principle, also be suppressed by a glycan substituent at the phenolic hydroxyl group. This would indicate that Tyr10 glycosylation is beneficial and that the sialylated O-glycans might prevent the nitration, and thereby the subsequent precipitation, of Aβ peptides. However, further studies are required to verify these speculations.

In mammals, carbohydrate modification of the Tyr hydroxyl group was previously only known for the glucosylation of glycogenin [172, 173]. However, the identification of a HexNAc-O-Tyr containing glycopeptide, originating from Nucleobindin 2, was also identified in the recent study of Clausen and co-workers [109]. Based on the lectin specificity of the ZFN approach, these authors concluded that the glycan identity is most likely GalNAc-O-Tyr. This may also apply to the Aβ glycopeptides, i.e. the HexNAc identity might be GalNAc-O-Tyr. However, further studies are required to elucidate the exact structure of these sialylated O-glycans.

Urinary glycoproteomics

The sialic acid capture-and-release strategy was also used for the enrichment of urinary N- and O-glycopeptides (paper **III**). Initially, the enrichment procedure was hampered by the presence of metabolic waste products in the urine samples (section 1.4). These metabolic waste products, responsible for quenching the mild periodate oxidation of sialic acids, were eventually removed by dialysis but only after the addition of 1.5% SDS. These results demonstrated that, although sialic acids may be selectively oxidized, an additional requirement, apart from non-modified C7-C9 hydroxyls, is the absence of metabolic waste products or other interfering compounds. Nevertheless, this issue might be avoided by efficient protein purification procedures.

Following capture-and-release enrichment, 58 N- and 63 O-glycopeptides were identified by CID and ECD experiments. In contrast to O-glycopeptides from CSF, extensive glycan heterogeneity was observed for specific O-glycopeptides enriched from the urine samples (Fig. 15). However, the MS analysis of the heterogeneous O-glycopeptides proved to be difficult for several reasons, including: i) the inability of B/Y-type ions to differentiate core 2-like glycans from elongated (linear) core 1-like glycans and ii) the

inability of Y-type fragments to differentiate peptides occupied by a single core 2-like glycan from peptides occupied by two individual core 1-like glycans. However, by combining the CID-MS^n data with the ECD data for the same precursor ions, complementary information could be gathered and used to differentiate these types of heterogeneous O-glycosylations. Although urinary O-glycopeptides were carefully analyzed, Tyr glycosylated peptides were not identified. These results suggest that complex tyrosine glycosylation is rare, and possibly more tissue specific, than mucin-type O-glycosylation on Ser/Thr residues.

6 CONCLUSION

The overall goal of this thesis was to develop glycoproteomic techniques for characterization of N- and O-linked glycoproteins in complex biological samples.

Due to the structural complexity of glycoproteins no single analytical technique is available for a complete structural characterization of any glycoconjugate. Thus, some kind of reductionistic approach is usually employed to simplify, automate and interpret the analyses. The basic analytical set-up is usually one of several variants of chromatography coupled to mass spectrometry, due to the excellent capacity for separation, speed, sensitivity, characteristic fragmentation and high mass accuracy of low to high molecular weight compounds. Traditionally, glycobiologists release from glycoproteins the glycans to fully characterize the glycome, whereas protein chemists and other biologists release the glycans to fully characterize the proteome. The ambition of this thesis has been to start bridging the two approaches in order to facilitate for both fields to incorporate knowledge from the other, simply asking the question which glycan is attached to which amino acid residue in a protein. Additionally, in order to cope with the large dynamic width of protein concentrations as well as their structural complexities in biological samples one or several efficient procedures are needed for enrichment or simplification of the mixture of glycoprotein isoforms subjected to analysis. Finally, the vast amount of data generated from the LC-MS/MS analyses must be handled, interpreted and made available for others to use. These aspects of glycoproteomic approaches are the cornerstones of this thesis.

Terminal sialic acids are characteristically found on many oligosaccharides of glycoproteins and play important roles for the processing, halftime and biological functions of sialylated glycoproteins. From this standpoint and the relative ease to specifically oxidize the side chain of sialic acid residues and covalently couple them to hydrazide (beads), perform trypsin digestion of solid phase attached proteins and finally cleave the characteristically acid labile sialic acid glycosidic bond by mild acid treatment, we introduced a procedure for global characterization of sialylated glycoproteins. This sialic acid capture-and-release protocol was shown to be a successful approach for the efficient analysis of N- and O-glycopeptides derived from glycoproteins/glycopeptides in human cerebrospinal fluid and urine. The urine samples however, could not initially be treated simply according to the original protocol for oxidation of sialic acids, but needed extensive dialysis to

remove reducing agents. By also introducing a pretreatment step using PNGase F for enzymatic removal of N-linked glycans, prior to sialic acid oxidation, focus was put specifically onto the O-glycoproteome. Altogether 66 glycoproteins and 131 glycosylation sites (36 N-linked and 95 O-linked) were identified from the CSF samples. The corresponding figures for urine samples were 53 glycoproteins and 82 glycosylation sites (25 N-linked and 57 O-linked). The majority (45) of the O-glycosylation sites were previously unknown.

Initially we analyzed all LC-MS/MS spectra manually but ultimately we were able to automate the procedures to export CID-MS3 peptide fragmentation peak lists together with the calculated peptide precursor mass into a Mascot file in order to simplify the identification of significant peptides. This automated analysis of spectra of O-glycopeptides should make our sialic acid-capture-and-release method more attractive also for other users. With increasingly more and more O-glycosylation sites being identified on human glycoproteins an opportunity was given to statistically analyze the residues surrounding the glycosylated Ser/Thr residues in order to identify possible glycosylation motifs. Indeed, we found that Thr/Ser/Pro residues in S/T-X-X-P (45%), P-S/T (35%) and S/T-P (25%) glycosylation motifs were predominant features of the O-glycopeptides from CSF.

Through the sialic acid targeted O-glycoproteomic analysis of CSF we identified glycopeptides in the vicinity of the β-secretase cleavage site of the amyloid precursor protein (APP). By specifically analyzing APP/Aβ fragments, immunoprecipitated from CSF, an in-depth glyco/proteomic analysis of APP was performed. This study confirmed the findings of a series of short Aβ1-X peptides but also revealed two completely novel series of O-glycopeptides, one being APP/AβX-15 glycopeptides carrying up to 4 mono- or disialylated core 1 like O-glycans linked to Thr residues and the other being an Aβ1-X series of glycopeptides carrying one core l-like glycan with up to three sialic acid residues and uniquely linked to a specific tyrosine residue. Interestingly, in a small number of CSF samples from patients with Alzheimer disease or with mild cognitive impairment, the concentration of the latter series of glycopeptides was enriched relative to that of the Aβ peptides found in CSF. We were not able to identify any glycosylated form of the amyloidogenic Aβ1-42 in any of the samples. The pathobiological significance as well as the diagnostic potential of these findings still remains to be shown.

7 FUTURE PERSPECTIVES

Although the work presented in this thesis has opened up some new strategies for future glycoproteomic studies there are still many technical challenges to tackle.

Obvious limitations of our sialic acid capture-and-release protocol are e.g. that we do not enrich for non-sialylated structures and that we have so far cleaved off the sialic acids in the release protocol, thus loosing information on numbers of sialic acids and their positions in the enriched glycans. One strategy would be to refine the release mechanism to enable analysis of sialylated O-glycopeptides and thus address the issue of sialic acid heterogeneity in various glycoprotein isoforms. Additionally, the chromatographic systems that we have used are essentially taken only from standard proteomics LC-MS/MS protocols and have not been optimized for separation according to glycan but rather to peptide characteristics (i.e. hydrophobicity). The efficiency of protease digestion may also be challenged by using other enzymes to release glycopeptides of suitable length and composition. Finally there is a need to develop the LC-MS/MS analyses so that ions in the low mass region - or any ions of specific interest - are included in successive fragmentation cycles to yield optimal information on e.g. linkage positions and possibly monosaccharide identities. There is also a need to increase the sensitivity of the analyses to single-cell levels, to automate the procedures and to go from relative to absolute quantifications using isotopically labeled standards of glycopeptides. Finally, the bioinformatics handling of data should be improved from tedious manual interpretations, as used in the beginning of this thesis, to automated searches in relevant databases. Interactive systems for web based interpretation of MS data files should help in building reference based knowledge in the establishment of the field of glycoproteomics.

Although limitations exist with the technology presented in this thesis there are certainly applications that call for activity. So far we have analyzed human CSF and urine samples but for making relevant comparisons with data already generated, it would certainly be interesting to analyze human plasma samples. In search for biomarkers and future diagnostic purposes plasma is generally considered the standard system for sampling and assaying. LC-MS/MS analyses are well established in clinical laboratories but generally too slow and costly for standard assays and are usually complemented with antibody based easy access assays. However, there are no principal restrictions in the methodology presented not to analyze glycoproteins of

lysed cells and cell membranes. Such preparation may be taken from blood cells or from cell cultures. Of particular interest for us is to follow up our recent findings on the glycosylation of tyrosine in APP producing cells, of O-mannosyl linked glycosylation of α-dystroglycan in differentiated muscle and nerve cells, as well as the importance of glycosylation of viral glycoproteins in cells facilitating virus propagation.

ACKNOWLEDGEMENT

There are many persons that have, in one way or another, supported me during my time as a PhD student and a proper acknowledgment would require a lot more than the few pages left in this book. I would like to express my deepest gratitude to all of you and I would especially like to thank…

Göran Larson, first of all for accepting me as your student, for your excellent tutorship and guidance in the field of research, for your never-ending enthusiasm and support, for taking your time to answer (occasionally stupid) questions, for working days and (late) nights with abstracts and manuscripts. I can go on and on but I will stop here by saying that all students should be supervised by someone like you.

Jonas Nilsson, for co-supervising me and teaching me all you know about chemistry and MS, for helping me to "crack" enigmatic spectra and for the good times in Texas, Puerto Rico and all of the other places we've been to together. Working with you has been a great experience.

Camilla Hesse, for co-supervising me and for introducing me into the wonderful world of platelets, for your enthusiasm and support, for fruitful discussions and for always being helpful.

The **GL**ycobiology group: **Ammi**, for sharing your wealth of knowledge and taking care of me in the lab, my room-mates **Gustaf**, **Johanna**, **Waqas** and **Alejandro**, my colleagues **Inger** and **Angelika**, for the good times we had and for all scientific and non-scientific discussions. **Gitti**. **Eva**.

Ulla Rüetschi, for all your help with MS instruments and analyses.

The Breimer group: **Lola**, **Angela** and **Mette**, for being helpful in the lab and for all the good times we had during lunch and coffee breaks. The Lindahl group: **Johan**, **Kristina**, **Narmin**, **Lars**.

Collaborators and co-authors: **Kaj Blennow, Henrik Zetterberg, Gunnar Brinkmalm, Erik Portelius, Ann Westman-Brinkmalm, Sigge Olofsson, Rickard Norden, Hakon Leffler, Cecilia Cederfur**

My half-time seminar committee: **Ola Hammarsten**, **Niklas Karlsson** and **Jan-Eric Månsson**

Karl-Anders Karlsson, for the inspirational scientific discussions.

Ulla Strandberg, for your help with administrative matters.

Susann Teneberg, for introducing me to the field of glycobiology and for all of the other things you have done for me. I owe you big-time. “Gänget”: **Marianne**, **Catharina**, **Lena** and **John**, for the good times we had. I’m looking forward to many more of our social gatherings.

The core facility people: **Elisabet**, **Petra**, **Carina, Sjoerd**, for analyzing our samples, and **Jörgen** (I still remember the day you asked: do you know what a sialic acid is?)

To my friends, for just being that. You know who you are.

Hassan abi and **Kazim** dağı, for inspiring me and shaping me into the person I am.

To my family, **Mom**, **Dad**, brother **Halim**, sister-in-law **Semiha**, **Arda**, **Aleyna**, **Najah**, **Zaineb**, for love and support during good and bad times. This thesis is dedicated to you.

To my fiancée **Sara**, for everything that you do for me.

REFERENCES

1. Marth, J.D., *A unified vision of the building blocks of life.* Nat Cell Biol, 2008. **10**(9): p. 1015-6.
2. Varki, A., J.D. Esko, and K.J. Colley, *Cellular Organization of Glycosylation*, in *Essentials of Glycobiology*, A. Varki, et al., Editors. 2009, Cold Spring Harbor Laboratory Press: New York. p. 36-61.
3. Gabius, H.J., *Glycans: bioactive signals decoded by lectins.* Biochem Soc Trans, 2008. **36**(Pt 6): p. 1491-6.
4. Varki, A., et al., *Discovery and Classification of Glycan-Binding Proteins*, in *Essentials of Glycobiology*, A. Varki, et al., Editors. 2009, Cold Spring Harbor Laboratory Press: New York. p. 375-386.
5. Rademacher, T.W., R.B. Parekh, and R.A. Dwek, *Glycobiology.* Annu Rev Biochem, 1988. **57**: p. 785-838.
6. McMurry, J., *Stereochemistry*, in *Organic Chemistry*. 2000, Brooks/Cole: Pacific Grove, CA. p. 306-354.
7. Varki, A. and R. Schauer, *Sialic Acids*, in *Essentials of Glycobiology*, A. Varki, et al., Editors. 2009, Cold Spring Harbor Laboratory Press: New York. p. 199-217.
8. Traving, C. and R. Schauer, *Structure, function and metabolism of sialic acids.* Cell Mol Life Sci, 1998. **54**(12): p. 1330-49.
9. Freeze, H.H. and A.D. Elbein, *Glycosylation Precursors*, in *Essentials of Glycobiology*, A. Varki, et al., Editors. 2009, Cold Spring Harbor Laboratory Press: New York. p. 47-61.
10. Bertozzi, C.R. and D. Rabuka, *Structural Basis of Glycan Diversity*, in *Essentials of Glycobiology*, A. Varki, et al., Editors. 2009, Cold Spring Harbor Laboratory Press: New York. p. 23-36.
11. Rini, J., J. Esko, and A. Varki, *Glycosyltransferases and Glycan-processing Enzymes*, in *Essentials of Glycobiology*, A. Varki, et al., Editors. 2009, Cold Spring Harbor Laboratory Press: New York. p. 63-73.
12. Schauer, R., *Biosynthesis and function of N- and O-substituted sialic acids.* Glycobiology, 1991. **1**(5): p. 449-52.
13. Hiraoka, N., et al., *A novel, high endothelial venule-specific sulfotransferase expresses 6-sulfo sialyl Lewis(x), an L-selectin ligand displayed by CD34.* Immunity, 1999. **11**(1): p. 79-89.
14. Schauer, R., et al., *Metabolism and role of O-acetylated sialic acids.* Adv Exp Med Biol, 2001. **491**: p. 325-42.
15. Yoshida-Moriguchi, T., et al., *O-mannosyl phosphorylation of alpha-dystroglycan is required for laminin binding.* Science, 2010. **327**(5961): p. 88-92.

16. Stanley, P. and R.D. Cummings, *Structures Common to Different Glycans*, in *Essentials of Glycobiology*, A. Varki, et al., Editors. 2009, Cold Spring Harbor Laboratory Press: New York. p. 175-198.
17. Spiro, R.G., *Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds.* Glycobiology, 2002. **12**(4): p. 43R-56R.
18. Marshall, R.D., *The nature and metabolism of the carbohydrate-peptide linkages of glycoproteins.* Biochem Soc Symp, 1974(40): p. 17-26.
19. Bause, E. and G. Legler, *The role of the hydroxy amino acid in the triplet sequence Asn-Xaa-Thr(Ser) for the N-glycosylation step during glycoprotein biosynthesis.* Biochem J, 1981. **195**(3): p. 639-44.
20. Trombetta, E.S. and A.J. Parodi, *Quality control and protein folding in the secretory pathway.* Annu Rev Cell Dev Biol, 2003. **19**: p. 649-76.
21. Ruddock, L.W. and M. Molinari, *N-glycan processing in ER quality control.* J Cell Sci, 2006. **119**(Pt 21): p. 4373-80.
22. Stanley, P., H. Schachter, and N. Taniguchi, *N-Glycans*, in *Essentials of Glycobiology*, A. Varki, et al., Editors. 2009, Cold Spring Harbor Laboratory Press: New York. p. 101-114.
23. Chu, C.S., et al., *Profile of native N-linked glycan structures from human serum using high performance liquid chromatography on a microfluidic chip and time-of-flight mass spectrometry.* Proteomics, 2009. **9**(7): p. 1939-51.
24. Hard, K., et al., *The Asn-linked carbohydrate chains of human Tamm-Horsfall glycoprotein of one male. Novel sulfated and novel N-acetylgalactosamine-containing N-linked carbohydrate chains.* Eur J Biochem, 1992. **209**(3): p. 895-915.
25. Gill, D.J., H. Clausen, and F. Bard, *Location, location, location: new insights into O-GalNAc protein glycosylation.* Trends Cell Biol, 2011. **21**(3): p. 149-58.
26. Van den Steen, P., et al., *Concepts and principles of O-linked glycosylation.* Crit Rev Biochem Mol Biol, 1998. **33**(3): p. 151-208.
27. Ten Hagen, K.G., T.A. Fritz, and L.A. Tabak, *All in the family: the UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases.* Glycobiology, 2003. **13**(1): p. 1R-16R.
28. Tabak, L.A., *The role of mucin-type O-glycans in eukaryotic development.* Semin Cell Dev Biol, 2010. **21**(6): p. 616-21.
29. Brockhausen, I., H. Schachter, and P. Stanley, *O-GalNAc Glycans*, in *Essentials of Glycobiology*, A. Varki, et al., Editors.

2009, Cold Spring Harbor Laboratory Press: New York. p. 115-127.
30. Ju, T., et al., *Cloning and expression of human core 1 beta1,3-galactosyltransferase.* J Biol Chem, 2002. **277**(1): p. 178-86.
31. Wang, Y., et al., *Cosmc is an essential chaperone for correct protein O-glycosylation.* Proc Natl Acad Sci U S A, 2010. **107**(20): p. 9228-33.
32. Bierhuizen, M.F. and M. Fukuda, *Expression cloning of a cDNA encoding UDP-GlcNAc:Gal beta 1-3-GalNAc-R (GlcNAc to GalNAc) beta 1-6GlcNAc transferase by gene transfer into CHO cells expressing polyoma large tumor antigen.* Proc Natl Acad Sci U S A, 1992. **89**(19): p. 9326-330.
33. Vavasseur, F., et al., *Synthesis of O-glycan core 3: characterization of UDP-GlcNAc: GalNAc-R beta 3-N-acetyl-glucosaminyltransferase activity from colonic mucosal tissues and lack of the activity in human cancer cell lines.* Glycobiology, 1995. **5**(3): p. 351-7.
34. Schwientek, T., et al., *Control of O-glycan branch formation. Molecular cloning and characterization of a novel thymus-associated core 2 beta1, 6-n-acetylglucosaminyltransferase.* J Biol Chem, 2000. **275**(15): p. 11106-13.
35. Iwai, T., et al., *Molecular cloning and characterization of a novel UDP-GlcNAc:GalNAc-peptide beta1,3-N-acetylglucosaminyltransferase (beta 3Gn-T6), an enzyme synthesizing the core 3 structure of O-glycans.* J Biol Chem, 2002. **277**(15): p. 12802-9.
36. Tian, E. and K.G. Ten Hagen, *Recent insights into the biological roles of mucin-type O-glycosylation.* Glycoconj J, 2009. **26**(3): p. 325-34.
37. Crocker, P.R., *Siglecs: sialic-acid-binding immunoglobulin-like lectins in cell-cell interactions and signalling.* Curr Opin Struct Biol, 2002. **12**(5): p. 609-15.
38. Crocker, P.R., J.C. Paulson, and A. Varki, *Siglecs and their roles in the immune system.* Nat Rev Immunol, 2007. **7**(4): p. 255-66.
39. Cummings, R.D., *Structure and function of the selectin ligand PSGL-1.* Braz J Med Biol Res, 1999. **32**(5): p. 519-28.
40. Lowe, J.B., *Glycan-dependent leukocyte adhesion and recruitment in inflammation.* Curr Opin Cell Biol, 2003. **15**(5): p. 531-8.
41. Hang, H.C. and C.R. Bertozzi, *The chemistry and biology of mucin-type O-linked glycosylation.* Bioorg Med Chem, 2005. **13**(17): p. 5021-34.
42. Juge, N., *Microbial adhesins to gastrointestinal mucus.* Trends Microbiol, 2011(doi:10.1016/j.tim.2011.10.001).

43. Semenov, A.G., et al., *Processing of pro-brain natriuretic peptide is suppressed by O-glycosylation in the region close to the cleavage site.* Clin Chem, 2009. **55**(3): p. 489-98.
44. May, P., et al., *Differential glycosylation regulates processing of lipoprotein receptors by gamma-secretase.* J Biol Chem, 2003. **278**(39): p. 37386-92.
45. Maryon, E.B., S.A. Molloy, and J.H. Kaplan, *O-linked glycosylation at threonine 27 protects the copper transporter hCTR1 from proteolytic cleavage in mammalian cells.* J Biol Chem, 2007. **282**(28): p. 20376-87.
46. Seidah, N.G., *What lies ahead for the proprotein convertases?* Ann N Y Acad Sci, 2011. **1220**: p. 149-61.
47. Kato, K., et al., *Polypeptide GalNAc-transferase T3 and familial tumoral calcinosis. Secretion of fibroblast growth factor 23 requires O-glycosylation.* J Biol Chem, 2006. **281**(27): p. 18370-7.
48. Schjoldager, K.T., et al., *O-glycosylation modulates proprotein convertase activation of angiopoietin-like protein 3: possible role of polypeptide GalNAc-transferase-2 in regulation of concentrations of plasma lipids.* J Biol Chem, 2010. **285**(47): p. 36293-303.
49. Gram Schjoldager, K.T., et al., *A systematic study of site-specific GalNAc-type O-glycosylation modulating proprotein convertase processing.* J Biol Chem, 2011. **286**(46): p. 40122-32.
50. Matsui, T., K. Titani, and T. Mizuochi, *Structures of the asparagine-linked oligosaccharide chains of human von Willebrand factor. Occurrence of blood group A, B, and H(O) structures.* J Biol Chem, 1992. **267**(13): p. 8723-31.
51. Podbielska, M., et al., *ABH blood group antigens in O-glycans of human glycophorin A.* Arch Biochem Biophys, 2004. **429**(2): p. 145-53.
52. Fredriksson, S.A., et al., *ABH blood group antigens in N-glycan of human glycophorin A.* Arch Biochem Biophys, 2010. **498**(2): p. 127-35.
53. Varki, A., *Biological roles of oligosaccharides: all of the theories are correct.* Glycobiology, 1993. **3**(2): p. 97-130.
54. Varki, A. and J.B. Lowe, *Biological Roles of Glycans*, in *Essentials of Glycobiology*, A. Varki, et al., Editors. 2009, Cold Spring Harbor Laboratory Press: New York. p. 75-88.
55. Dell, A. and H.R. Morris, *Glycoprotein structure determination by mass spectrometry.* Science, 2001. **291**(5512): p. 2351-6.
56. Karas, M. and F. Hillenkamp, *Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons.* Anal Chem, 1988. **60**(20): p. 2299-301.

57. Fenn, J.B., et al., *Electrospray ionization for mass spectrometry of large biomolecules.* Science, 1989. **246**(4926): p. 64-71.
58. Koppenaal, D.W., et al., *MS detectors.* Anal Chem, 2005. **77**(21): p. 418A-427A.
59. Taylor, G., *Disintegration of Water Drops in an Electric Field.* Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences, 1964. **280**(1382): p. 383-397.
60. Dole, M., et al., *Molecular Beams of Macroions.* The Journal of Chemical Physics, 1968. **49**(5): p. 2240-2249.
61. Meng, C.K., M. Mann, and J.B. Fenn, *Of protons or proteins.* Zeitschrift für Physik D Atoms, Molecules and Clusters, 1988. **10**(2): p. 361-368.
62. Wilm, M.S. and M. Mann, *Electrospray and Taylor-Cone theory, Dole's beam of macromolecules at last?* International Journal of Mass Spectrometry and Ion Processes, 1994. **136**(2-3): p. 167-180.
63. Grimm, R.L. and J.L. Beauchamp, *Evaporation and discharge dynamics of highly charged multicomponent droplets generated by electrospray ionization.* J Phys Chem A, 2010. **114**(3): p. 1411-9.
64. Wilm, M., *Principles of electrospray ionization.* Mol Cell Proteomics, 2011. **10**(7): p. M111 009407.
65. Iribarne, J.V. and B.A. Thomson, *On the evaporation of small ions from charged droplets.* The Journal of Chemical Physics, 1976. **64**(6): p. 2287-2294.
66. Thomson, B.A. and J.V. Iribarne, *Field induced ion evaporation from liquid surfaces at atmospheric pressure.* The Journal of Chemical Physics, 1979. **71**(11): p. 4451-4463.
67. Siuzdak, G., et al., *Mass spectrometry and viral analysis.* Chem Biol, 1996. **3**(1): p. 45-8.
68. Wilm, M. and M. Mann, *Analytical properties of the nanoelectrospray ion source.* Anal Chem, 1996. **68**(1): p. 1-8.
69. Comisarow, M.B. and A.G. Marshall, *Fourier transform ion cyclotron resonance spectroscopy.* Chemical Physics Letters, 1974. **25**(2): p. 282-283.
70. Marshall, A.G., C.L. Hendrickson, and G.S. Jackson, *Fourier transform ion cyclotron resonance mass spectrometry: a primer.* Mass Spectrom Rev, 1998. **17**(1): p. 1-35.
71. Comisarow, M.B. and A.G. Marshall, *The early development of Fourier transform ion cyclotron resonance (FT-ICR) spectroscopy.* J Mass Spectrom, 1996. **31**(6): p. 581-5.
72. Amster, I.J., *Fourier Transform Mass Spectrometry.* Journal of Mass Spectrometry, 1996. **31**(12): p. 1325-1337.

73. Easterling, M.L., T.H. Mize, and I.J. Amster, *Routine Part-per-Million Mass Accuracy for High- Mass Ions: Space-Charge Effects in MALDI FT-ICR.* Anal Chem, 1999. **71**(3): p. 624-32.
74. Meng, F., et al., *Detection and localization of protein modifications by high resolution tandem mass spectrometry.* Mass Spectrom Rev, 2005. **24**(2): p. 126-34.
75. He, F., C.L. Hendrickson, and A.G. Marshall, *Baseline mass resolution of peptide isobars: a record for molecular mass resolution.* Anal Chem, 2001. **73**(3): p. 647-50.
76. Schwartz, J.C., M.W. Senko, and J.E. Syka, *A two-dimensional quadrupole ion trap mass spectrometer.* J Am Soc Mass Spectrom, 2002. **13**(6): p. 659-69.
77. Syka, J.E., et al., *Novel linear quadrupole ion trap/FT mass spectrometer: performance characterization and use in the comparative analysis of histone H3 post-translational modifications.* J Proteome Res, 2004. **3**(3): p. 621-6.
78. Pandey, A. and M. Mann, *Proteomics to study genes and genomes.* Nature, 2000. **405**(6788): p. 837-46.
79. Aebersold, R. and M. Mann, *Mass spectrometry-based proteomics.* Nature, 2003. **422**(6928): p. 198-207.
80. Shen, Y., et al., *High-efficiency nanoscale liquid chromatography coupled on-line with mass spectrometry using nanoelectrospray ionization for proteomics.* Anal Chem, 2002. **74**(16): p. 4235-49.
81. Perkins, D.N., et al., *Probability-based protein identification by searching sequence databases using mass spectrometry data.* Electrophoresis, 1999. **20**(18): p. 3551-67.
82. Bogdanov, B. and R.D. Smith, *Proteomics by FTICR mass spectrometry: top down and bottom up.* Mass Spectrom Rev, 2005. **24**(2): p. 168-200.
83. Steen, H. and M. Mann, *The ABC's (and XYZ's) of peptide sequencing.* Nat Rev Mol Cell Biol, 2004. **5**(9): p. 699-711.
84. Shevchenko, A., et al., *Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels.* Anal Chem, 1996. **68**(5): p. 850-8.
85. Wilkins, M.R., et al., *From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis.* Biotechnology (N Y), 1996. **14**(1): p. 61-5.
86. Van den Bergh, G. and L. Arckens, *Recent advances in 2D electrophoresis: an array of possibilities.* Expert Rev Proteomics, 2005. **2**(2): p. 243-52.
87. Hoffert, J.D. and M.A. Knepper, *Taking aim at shotgun phosphoproteomics.* Anal Biochem, 2008. **375**(1): p. 1-10.
88. Thingholm, T.E., O.N. Jensen, and M.R. Larsen, *Analytical strategies for phosphoproteomics.* Proteomics, 2009. **9**(6): p. 1451-68.

89. Drake, R.R., et al., *Lectin capture strategies combined with mass spectrometry for the discovery of serum glycoprotein biomarkers.* Mol Cell Proteomics, 2006. **5**(10): p. 1957-67.
90. Madera, M., et al., *High-sensitivity profiling of glycoproteins from human blood serum through multiple-lectin affinity chromatography and liquid chromatography/tandem mass spectrometry.* J Chromatogr B Analyt Technol Biomed Life Sci, 2007. **845**(1): p. 121-37.
91. Bertozzi, C.R. and R. Sasisekharan, *Glycomics*, in *Essentials of Glycobiology*, A. Varki, et al., Editors. 2009, Cold Spring Harbor Laboratory Press: New York. p. 679-689.
92. Carlson, D.M., *Structures and immunochemical properties of oligosaccharides isolated from pig submaxillary mucins.* J Biol Chem, 1968. **243**(3): p. 616-26.
93. Fukuda, M., *Beta-elimination for release of O-GalNAc-linked oligosaccharides from glycoproteins and glycopeptides.* Curr Protoc Mol Biol, 2001. **Chapter 17**: p. Unit17 15B.
94. O'Neill, R.A., *Enzymatic release of oligosaccharides from glycoproteins for chromatographic and electrophoretic analysis.* J Chromatogr A, 1996. **720**(1-2): p. 201-15.
95. Harvey, D.J., *Structural determination of N-linked glycans by matrix-assisted laser desorption/ionization and electrospray ionization mass spectrometry.* Proteomics, 2005. **5**(7): p. 1774-86.
96. Tarentino, A.L., C.M. Gomez, and T.H. Plummer, Jr., *Deglycosylation of asparagine-linked glycans by peptide:N-glycosidase F.* Biochemistry, 1985. **24**(17): p. 4665-71.
97. Ruhaak, L.R., et al., *Glycan labeling strategies and their use in identification and quantification.* Anal Bioanal Chem, 2010. **397**(8): p. 3457-81.
98. Haslam, S.M., S.J. North, and A. Dell, *Mass spectrometric analysis of N- and O-glycosylation of tissues and cells.* Curr Opin Struct Biol, 2006. **16**(5): p. 584-91.
99. Geyer, H. and R. Geyer, *Strategies for analysis of glycoprotein glycosylation.* Biochim Biophys Acta, 2006. **1764**(12): p. 1853-69.
100. Wuhrer, M., et al., *Glycoproteomics based on tandem mass spectrometry of glycopeptides.* J Chromatogr B Analyt Technol Biomed Life Sci, 2007. **849**(1-2): p. 115-28.
101. Madera, M., Y. Mechref, and M.V. Novotny, *Combining lectin microcolumns with high-resolution separation techniques for enrichment of glycoproteins and glycopeptides.* Anal Chem, 2005. **77**(13): p. 4081-90.
102. Alvarez-Manilla, G., et al., *Tools for glycoproteomic analysis: size exclusion chromatography facilitates identification of tryptic*

glycopeptides with N-linked glycosylation sites. J Proteome Res, 2006. **5**(3): p. 701-8.
103. Thaysen-Andersen, M., et al., *Rapid and individual-specific glycoprofiling of the low abundance N-glycosylated protein tissue inhibitor of metalloproteinases-1.* Mol Cell Proteomics, 2007. **6**(4): p. 638-47.
104. Mysling, S., et al., *Utilizing ion-pairing hydrophilic interaction chromatography solid phase extraction for efficient glycopeptide enrichment in glycoproteomics.* Anal Chem, 2010. **82**(13): p. 5598-609.
105. Christiansen, M.N., et al., *Challenges of determining O-glycopeptide heterogeneity: a fungal glucanase model system.* Anal Chem, 2010. **82**(9): p. 3500-9.
106. Jensen, P.H., D. Kolarich, and N.H. Packer, *Mucin-type O-glycosylation--putting the pieces together.* FEBS J, 2010. **277**(1): p. 81-94.
107. Darula, Z. and K.F. Medzihradszky, *Affinity enrichment and characterization of mucin core-1 type glycopeptides from bovine serum.* Mol Cell Proteomics, 2009. **8**(11): p. 2515-26.
108. Kurogochi, M., et al., *Sialic acid-focused quantitative mouse serum glycoproteomics by multiple reaction monitoring assay.* Mol Cell Proteomics, 2010. **9**(11): p. 2354-68.
109. Steentoft, C., et al., *Mining the O-glycoproteome using zinc-finger nuclease-glycoengineered SimpleCell lines.* Nat Methods, 2011. **8**(11): p. 977-82.
110. Pan, S., et al., *Mass spectrometry based glycoproteomics--from a proteomics perspective.* Mol Cell Proteomics, 2011. **10**(1): p. R110 003251.
111. Ohtsubo, K. and J.D. Marth, *Glycosylation in cellular mechanisms of health and disease.* Cell, 2006. **126**(5): p. 855-67.
112. Kim, E.H. and D.E. Misek, *Glycoproteomics-based identification of cancer biomarkers.* Int J Proteomics, 2011. **2011**: p. 601937.
113. Hwang, H., et al., *Glycoproteomics in neurodegenerative diseases.* Mass Spectrom Rev, 2010. **29**(1): p. 79-125.
114. Johanson, C.E., *Choroid Plexus–Cerebrospinal Fluid Circulatory Dynamics: Impact on Brain Growth, Metabolism, and Repair*, in *Neuroscience in Medicine*, P.M. Conn, Editor. 2008, Humana Press. p. 173-200.
115. Brown, P.D., et al., *Molecular mechanisms of cerebrospinal fluid production.* Neuroscience, 2004. **129**(4): p. 957-70.
116. Heimer, L., *The human brain and spinal cord*. 2nd ed. 1995, New York: Springer-Verlag.
117. Zhang, J., et al., *Quantitative proteomic analysis of age-related changes in human cerebrospinal fluid.* Neurobiol Aging, 2005. **26**(2): p. 207-27.

118. Schutzer, S.E., et al., *Establishing the proteome of normal human cerebrospinal fluid.* PLoS One, 2010. **5**(6): p. e10980.
119. Alzheimer, A., et al., *An English translation of Alzheimer's 1907 paper, "Uber eine eigenartige Erkankung der Hirnrinde".* Clin Anat, 1995. **8**(6): p. 429-31.
120. Blennow, K., M.J. de Leon, and H. Zetterberg, *Alzheimer's disease.* Lancet, 2006. **368**(9533): p. 387-403.
121. Alonso, A., et al., *Hyperphosphorylation induces self-assembly of tau into tangles of paired helical filaments/straight filaments.* Proc Natl Acad Sci U S A, 2001. **98**(12): p. 6923-8.
122. Masters, C.L., et al., *Amyloid plaque core protein in Alzheimer disease and Down syndrome.* Proc Natl Acad Sci U S A, 1985. **82**(12): p. 4245-9.
123. Kang, J., et al., *The precursor of Alzheimer's disease amyloid A4 protein resembles a cell-surface receptor.* Nature, 1987. **325**(6106): p. 733-6.
124. Glenner, G.G. and C.W. Wong, *Alzheimer's disease: initial report of the purification and characterization of a novel cerebrovascular amyloid protein.* Biochem Biophys Res Commun, 1984. **120**(3): p. 885-90.
125. Haass, C., et al., *Amyloid beta-peptide is produced by cultured cells during normal metabolism.* Nature, 1992. **359**(6393): p. 322-5.
126. Goldgaber, D., et al., *Characterization and chromosomal localization of a cDNA encoding brain amyloid of Alzheimer's disease.* Science, 1987. **235**(4791): p. 877-80.
127. Lammich, S., et al., *Constitutive and regulated alpha-secretase cleavage of Alzheimer's amyloid precursor protein by a disintegrin metalloprotease.* Proc Natl Acad Sci U S A, 1999. **96**(7): p. 3922-7.
128. Vassar, R., et al., *Beta-secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BACE.* Science, 1999. **286**(5440): p. 735-41.
129. Wakabayashi, T., T. Iwatsubo, and B. Strooper, *The Biology of the Presenilin Complexes*, in *Alzheimer's disease: advances in genetics, molecular, and cellular biology*, S.S. Sisodia and R.E. Tanzi, Editors. 2007, Springer US. p. 35-58.
130. Portelius, E., et al., *Determination of beta-amyloid peptide signatures in cerebrospinal fluid using immunoprecipitation-mass spectrometry.* J Proteome Res, 2006. **5**(4): p. 1010-6.
131. Portelius, E., et al., *Characterization of amyloid beta peptides in cerebrospinal fluid by an automated immunoprecipitation procedure followed by mass spectrometry.* J Proteome Res, 2007. **6**(11): p. 4433-9.

132. Portelius, E., et al., *Identification of novel APP/Abeta isoforms in human cerebrospinal fluid.* Neurodegener Dis, 2009. **6**(3): p. 87-94.
133. Portelius, E., et al., *A novel pathway for amyloid precursor protein processing.* Neurobiol Aging, 2011. **32**(6): p. 1090-8.
134. Portelius, E., et al., *Identification of novel N-terminal fragments of amyloid precursor protein in cerebrospinal fluid.* Exp Neurol, 2010. **223**(2): p. 351-8.
135. Hardy, J.A. and G.A. Higgins, *Alzheimer's disease: the amyloid cascade hypothesis.* Science, 1992. **256**(5054): p. 184-5.
136. Selkoe, D.J., *The molecular pathology of Alzheimer's disease.* Neuron, 1991. **6**(4): p. 487-98.
137. First, M.R., *Renal Function*, in *Clinical chemistry: theory, analysis, and correlation*, L. Kaplan, Pesce, AJ., Editor. 1984, C.V. Mosby Co.: St. Louis, MO. p. 403-419.
138. Jefferson, J.A., S.J. Shankland, and R.H. Pichler, *Proteinuria in diabetic kidney disease: a mechanistic viewpoint.* Kidney Int, 2008. **74**(1): p. 22-36.
139. Jarad, G. and J.H. Miner, *Update on the glomerular filtration barrier.* Curr Opin Nephrol Hypertens, 2009. **18**(3): p. 226-32.
140. Patrakka, J. and K. Tryggvason, *Molecular make-up of the glomerular filtration barrier.* Biochem Biophys Res Commun, 2010. **396**(1): p. 164-9.
141. Haraldsson, B., J. Nystrom, and W.M. Deen, *Properties of the glomerular barrier and mechanisms of proteinuria.* Physiol Rev, 2008. **88**(2): p. 451-87.
142. Pisitkun, T., R.F. Shen, and M.A. Knepper, *Identification and proteomic profiling of exosomes in human urine.* Proc Natl Acad Sci U S A, 2004. **101**(36): p. 13368-73.
143. Thongboonkerd, V. and P. Malasit, *Renal and urinary proteomics: current applications and challenges.* Proteomics, 2005. **5**(4): p. 1033-42.
144. Heine, G., M. Raida, and W.G. Forssmann, *Mapping of peptides and protein fragments in human urine using liquid chromatography-mass spectrometry.* J Chromatogr A, 1997. **776**(1): p. 117-24.
145. Spahr, C.S., et al., *Towards defining the urinary proteome using liquid chromatography-tandem mass spectrometry. I. Profiling an unfractionated tryptic digest.* Proteomics, 2001. **1**(1): p. 93-107.
146. Pieper, R., et al., *Characterization of the human urinary proteome: a method for high-resolution display of urinary proteins on two-dimensional electrophoresis gels with a yield of nearly 1400 distinct protein spots.* Proteomics, 2004. **4**(4): p. 1159-74.

147. Castagna, A., et al., *Exploring the hidden human urinary proteome via ligand library beads.* J Proteome Res, 2005. **4**(6): p. 1917-30.
148. Adachi, J., et al., *The human urinary proteome contains more than 1500 proteins, including a large proportion of membrane proteins.* Genome Biol, 2006. **7**(9): p. R80.
149. Kentsis, A., et al., *Urine proteomics for profiling of human disease using high accuracy mass spectrometry.* Proteomics Clin Appl, 2009. **3**(9): p. 1052-1061.
150. Nagaraj, N. and M. Mann, *Quantitative analysis of the intra- and inter-individual variability of the normal urinary proteome.* J Proteome Res, 2011. **10**(2): p. 637-45.
151. Pejcic, M., S. Stojnev, and V. Stefanovic, *Urinary proteomics--a tool for biomarker discovery.* Ren Fail, 2010. **32**(2): p. 259-68.
152. Thongboonkerd, V., *Urinary proteomics: towards biomarker discovery, diagnostics and prognostics.* Mol Biosyst, 2008. **4**(8): p. 810-5.
153. Gahmberg, C.G. and L.C. Andersson, *Selective radioactive labeling of cell surface sialoglycoproteins by periodate-tritiated borohydride.* J Biol Chem, 1977. **252**(16): p. 5888-94.
154. Van Lenten, L. and G. Ashwell, *Studies on the chemical and enzymatic modification of glycoproteins. A general method for the tritiation of sialic acid-containing glycoproteins.* J Biol Chem, 1971. **246**(6): p. 1889-94.
155. Rafelson, M.E., Jr., H. Clauser, and J. Legault-Demare, *Removal of sialic acid from serum gonadotropin by acidic and enzymic hydrolysis.* Biochim Biophys Acta, 1961. **47**: p. 406-7.
156. Zhang, H., et al., *Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry.* Nat Biotechnol, 2003. **21**(6): p. 660-6.
157. Roepstorff, P. and J. Fohlman, *Proposal for a common nomenclature for sequence ions in mass spectra of peptides.* Biomed Mass Spectrom, 1984. **11**(11): p. 601.
158. Huddleston, M.J., M.F. Bean, and S.A. Carr, *Collisional fragmentation of glycopeptides by electrospray ionization LC/MS and LC/MS/MS: methods for selective detection of glycopeptides in protein digests.* Anal Chem, 1993. **65**(7): p. 877-84.
159. Domon, B. and C.E. Costello, *A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates.* Glycoconjugate Journal, 1988. **5**(4): p. 397-409.
160. Mirgorodskaya, E., P. Roepstorff, and R.A. Zubarev, *Localization of O-glycosylation sites in peptides by electron capture dissociation in a Fourier transform mass spectrometer.* Anal Chem, 1999. **71**(20): p. 4431-6.

161. Dodds, E.D., et al., *Analytical performance of immobilized pronase for glycopeptide footprinting and implications for surpassing reductionist glycoproteomics.* J Proteome Res, 2009. **8**(2): p. 502-12.
162. Annesley, T.M., *Ion suppression in mass spectrometry.* Clin Chem, 2003. **49**(7): p. 1041-4.
163. Nilsson, J., G. Larson, and A. Grahn, *Characterization of site-specific O-glycan structures within the mucin-like domain of alpha-dystroglycan from human skeletal muscle.* Glycobiology, 2010. **20**(9): p. 1160-9.
164. Kurogochi, M., et al., *Reverse glycoblotting allows rapid-enrichment glycoproteomics of biopharmaceuticals and disease-related biomarkers.* Angew Chem Int Ed Engl, 2007. **46**(46): p. 8808-13.
165. Li, A., et al., *Chemical cleavage at aspartyl residues for protein identification.* Anal Chem, 2001. **73**(22): p. 5395-402.
166. Zhang, Z. and W.J. Henzel, *Signal peptide prediction based on analysis of experimentally verified cleavage sites.* Protein Sci, 2004. **13**(10): p. 2819-24.
167. Wilson, I.B., Y. Gavel, and G. von Heijne, *Amino acid distributions around O-linked glycosylation sites.* Biochem J, 1991. **275 (Pt 2)**: p. 529-34.
168. Gupta, R., et al., *O-GLYCBASE version 4.0: a revised database of O-glycosylated proteins.* Nucleic Acids Res, 1999. **27**(1): p. 370-2.
169. Jankowsky, J.L., et al., *Rodent A beta modulates the solubility and distribution of amyloid deposits in transgenic mice.* J Biol Chem, 2007. **282**(31): p. 22707-20.
170. Kummer, M.P., et al., *Nitration of tyrosine 10 critically enhances amyloid beta aggregation and plaque formation.* Neuron, 2011. **71**(5): p. 833-44.
171. McMurry, J., *Benzene and Aromaticity*, in *Organic Chemistry*. 2000, Brooks/Cole: Pacific Grove, CA. p. 559-591.
172. Aon, M.A. and J.A. Curtino, *Protein-bound glycogen is linked to tyrosine residues.* Biochem J, 1985. **229**(1): p. 269-72.
173. Rodriguez, I.R. and W.J. Whelan, *A novel glycosyl-amino acid linkage: rabbit-muscle glycogen is covalently linked to a protein via tyrosine.* Biochem Biophys Res Commun, 1985. **132**(2): p. 829-36.