



MASTERARBEIT

Titel der Masterarbeit

Targeted glycoproteomics for monitoring changes in
glycosylation upon cell activation

verfasst von

Dietmar Hammerschmid, Bakk. rer. nat.

angestrebter akademischer Grad

Master of Science (MSc)

Wien, December 2014

Studienkennzahl lt. Studienblatt:

A 066 863

Studienrichtung lt. Studienblatt:

Masterstudium Biologische Chemie

Betreut von:

Univ.-Prof. Mag. Dr. Andreas Rizzi

I. CONTENTS

I. CONTENTS	3
II. ACKNOWLEDGEMENTS	6
III. LIST OF ABBREVIATIONS	8
IV. I - FUNDAMENTAL ASPECTS AND BACKGROUND	11
1. PROTEIN GLYCOSYLATION	11
1.1 N-linked Glycosylation	12
1.1.1 N-linked Glycan Biosynthesis	14
1.1.2 Biological Roles of Glycans	18
1.1.3 Glycoproteomics in Disease Research	19
2. ENRICHMENT STRATEGIES FOR PROTEINS AND GLYCOPROTEINS	21
2.1 Principle of Affinity Chromatography	21
2.1.1 Affinity Enrichment in Step Elution Mode	22
2.1.2 Affinity Depletion	23
2.2 Specific Enrichment of Glycoproteins and Glycopeptides	23
2.2.1 Lectin Affinity Chromatography (LAC)	24
3. MASS SPECTROMETRY BASED PROTEOME ANALYSIS	26
3.1 Electrospray Ionization (ESI)	26
3.1.1 Construction of ESI Source	27
3.1.2 Ionization Mechanism	28
3.2 Orbitrap Technology	29
3.2.1 Construction of Orbitrap Mass Analyzer	30
3.2.2 Theoretical Background	31
3.3 Tandem Mass Spectrometry (MS/MS)	32
3.3.1 CID Fragmentation	32
3.3.2 HCD Fragmentation	35
3.3.3 ETD Fragmentation	37
V. II - EXPERIMENTAL WORK	40
1. INTRODUCTION	40

2. MATERIALS AND METHODS	43
2.1 Samples, Chemicals and Instruments	43
2.1.1 Samples	43
2.1.2 Chemicals (Buffers and Solution)	43
2.1.3 Instruments	47
2.2 Methods	47
2.2.1 Cell Culture	47
2.2.2 SDS-PAGE for subsequent Shotgun Analysis	48
2.2.3 Silver staining	49
2.2.4 Reduction, Alkylation, and Trypsin Digest	49
2.2.5 MS Analysis	51
3. RESULTS AND DISCUSSION	55
3.1 Glycoprotein Identification in SN of MCF-7 Cells	55
3.2 Site specific <i>N</i>-Glycan structures in Thrombospondin-1	57
3.3 ETD Spectrum of Glycopeptides	66
3.4 Relative Quantification of different Glycopeptides	68
4. CONCLUSION	70
5. SUPPLEMENTAL MATERIAL	72
VI. REFERENCES	76
VII. ABSTRACT	85
VIII. ZUSAMMENFASSUNG (ABSTRACT GERMAN)	86
IX. CURRICULUM VITAE	88

II. ACKNOWLEDGEMENTS

First of all I am really thankful to Prof. Rizzi for giving me the possibility to be part of his research group during my experimental time and for his excellent supervision of my Master thesis on an interesting and challenging field of study. I am also grateful to Claudia Michael for introducing me into the laboratory and for helping me to build up my experiments. And I would like to thank Siniša Sic and Nikolaus Voulgaris, the other members of the research group, for the pleasant working atmosphere during my practical work.

Furthermore, I want to thank the team of the Mass Spectrometry Centre especially to Anna Fabisikova for introducing me into the LTQ Orbitrap Velos instrument. Moreover I am also thankful to Prof. Gerner and his group for the working cooperation and for the cell samples which were prepared by Andrea Bileck.

I would like to express my honest gratefulness to my family for their support over all those years of my study. Without their encouragement it wouldn't had been possible to achieve this graduation.

And at the end thank you to all my friends and colleagues for being with me all this time.

III. LIST OF ABBREVIATIONS

▪ ACN	Acetonitrile
▪ AD	Alzheimer disease
▪ AGC	Active gain control
▪ AGP	Alpha-1-acid glycoprotein
▪ APP	Amyloid precursor protein
▪ APS	Ammonium persulphate
▪ BPC	Base peak chromatogram
▪ CHAPS	3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate
▪ CID	Collision induced dissociation
▪ Con A	Concanavalin A
▪ DC	Direct current
▪ DMEM	Dulbecco modified Eagle's minimal essential medium
▪ DTT	Dithiothreitol
▪ ECD	Electron-capture dissociation
▪ ECM	Extracellular matrix
▪ ER	Endoplasmic reticulum
▪ ESI	Electrospray ionization
▪ ETD	Electron transfer dissociation
▪ FA	Formic acid
▪ FCS	Fetal calf serum
▪ FT	Fourier transform
▪ FTICR	Fourier transform ion cyclotron resonance
▪ GDP	Guanosine diphosphate
▪ HCD	Higher-energy collisional dissociation
▪ Hex	Hexose
▪ HexNAc	<i>N</i> -Acetylhexoseamine
▪ HPLC	High-performance liquid chromatography
▪ IAA	Iodoacetamide
▪ IgG	Immunoglobulin G
▪ LAC	Lectin Affinity Chromatography
▪ LLO	Lipid linked oligosaccharide

▪ LTQ	Linear trap quadrupole
▪ MCF-7	Michigan Cancer Foundation – 7 (Breast cancer cell line)
▪ MS	Mass spectrometry
▪ MS/MS	Tandem mass spectrometry
▪ Neu5Ac	<i>N</i> -Acetylneuraminic acid
▪ Neu5Gc	<i>N</i> -Glycolylneuraminic acid
▪ OST	Oligosaccharyltransferase
▪ PBS	Phosphate buffered saline
▪ PGC	Porous graphitized carbon
▪ PTM	Posttranslational modification
▪ RF	Radio frequency
▪ RP	Reversed phase
▪ SDS	Sodium dodecyl sulphate
▪ SN	Supernatant
▪ SNA	<i>Sambucus nigra</i> agglutinin
▪ TEMED	N,N,N',N' – tetramethylethylenediamine
▪ TOF	Time-of-flight
▪ TSP-1	Thrombospondin-1
▪ UDP	Uridine diphosphate

IV. I - FUNDAMENTAL ASPECTS and BACK-GROUND

1. Protein Glycosylation

Glycosylation is a common post-translational modification (PTM) by which a sugar is covalently attached to a functional group of a protein. Different monosaccharide units and only a few amino acids have been identified to be involved in the formation of glycoproteins leading to an increased diversity in the proteome. Protein glycosylation differs greatly among cell types and species. The attached oligosaccharides are believed to be involved in many different functions of the proteins within the cellular pathways.

11

Examinations have shown that 13 different monosaccharides and eight amino acids can be involved in forming glycopeptide bonds. Considering all possible combinations lead to at least 41 linkages able to occur in glycoproteins. Depending on the sugar and amino acid participating glycopeptide linkages, glycosylation can be arranged in five distinct groups.^[1]

N-linked glycosylation

N-glycosylation represents the most widely distributed sugar-amino acid linkage where the glycan is attached to the nitrogen of asparagine needing a certain consensus sequence (Asn-Xaa-Ser/Thr).^[2]

O-linked glycosylation

In O-linked glycosylation the glycan is commonly attached to the hydroxyl oxygen of the amino acids serine, threonine, and more rarely tyrosine, hydroxyproline, and hydroxylysine. In these compounds no consensus sequence is required.^[3]

C-linked glycosylation

Unlike the *N*- and *O*-linked glycosylation this linkage involves no functional group of an amino acid. C-linked glycans represent a rare form of glycosylation by which an α -mannosyl residue is attached to the C-2 of tryptophan. This type of glycosylation was first discovered in human RNase U_s.^[4]

Phosphoglycosylation

Phospho-glycans represent another distinct type of glycosylation where oligosaccharides are attached to phospho-serine in a polypeptide chain via a phosphodiester linkage.^[5]

Glypiation

A further form of glycosylation is the attachment of a glycosylphosphatidylinositol (GPI) anchor to proteins for attaching them to cell membranes. The carbohydrate-protein connection consists of mannoses linked to phosphoethanolamine, which is attached to the C-term of a protein. The GPI anchor has a phospholipid tail responsible for anchoring the protein to the membrane.^[6]

1.1 N-linked Glycosylation

N-linked glycosylation denotes a protein modification where *N*-glycans are covalently attached to asparagine (Asn) residues by an *N*-glycosidic bond. Early experiments on this field have shown that not every Asn is able to carry an *N*-glycan. The minimal consensus sequence consists of an asparagine followed by any amino acid except proline and ends with serine or threonine (Asn-Xaa-Ser/Thr).^[7] It has been proposed that the hydroxyl group of Ser or Thr is necessary for delivering a hydrogen-bond donor function in enzyme binding and in oligosaccharide transfer. Proline in Xaa-position of the consensus sequence might not be able to stabilize a loop conformation which is required for catalytic interaction between asparagine and the hydroxy amino acid.^[8] Nevertheless, transfer of an *N*-glycan to the Asn-Xaa-Ser/Thr sequon does not always happen, due to conformational constraints during glycoprotein folding. Recent studies have shown that in certain proteins (e.g.

immunoglobulin G (IgG) antibodies) glycosylation can also occur in modified consensus sequences.^[9] Beyond the Asn-Xaa-Ser/Thr sequon, Asn-Xaa-Cys motif has also been identified as a possible consensus sequence for *N*-glycosylation.^[10] So probably sequence requirements are not that strict than previously postulated.

The β -glycosidic bond between *N*-acetylglucosamine (GlcNAc) and asparagine occurring in *N*-linked glycosylation is pictured in Figure 1.

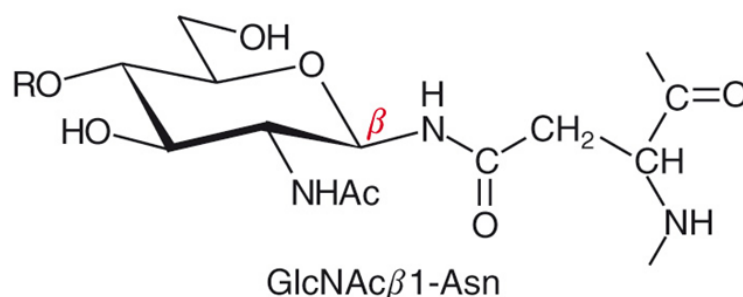


Figure 1: β -glycosidic bond between *N*-acetylglucosamine and amide nitrogen of asparagine.
R: oligosaccharyl residue. Picture taken from *Bioanalytik*, Springer-Verlag, Berlin Heidelberg, 2012.

A great number of different glycan structures have been observed in both eukaryotes and prokaryotes. All of these identified *N*-glycans share a common pentasaccharide core consisting of two GlcNAc and three mannose units. Depending on the extension of the core, *N*-glycans can be classified into complex, hybrid, and high-mannose type (Figure 2). In complex type, so called antennas consisting of different sugar units (N-acetylglucosamine, galactose, and sialinic acid), are attached to the core structure. The high-mannose type carries only mannose residues beside the common core structure. The hybrid type is a combination of both complex as well as high-mannose type, where the mannose residues are attached to the Man α 1-6 arm and the antennas are linked to the Man α 1-3 arm of the core.^[11]

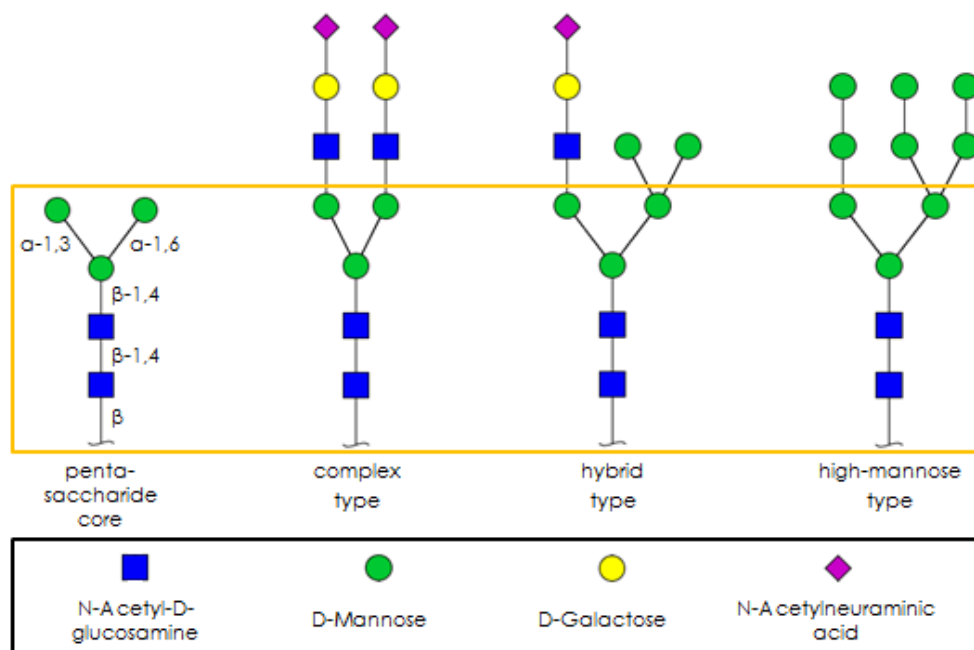


Figure 2: Different types of *N*-linked oligosaccharide structures with common pentasaccharide core (highlighted in the orange box). Types of linkages between monosaccharide units within the core are indicated.

1.1.1 *N*-linked Glycan Biosynthesis

The pathway of *N*-glycan biosynthesis can be separated into three major steps. The process starts with the synthesis of an oligosaccharide linked to dolichol precursor. The second step is characterised by *en bloc* transfer of the oligosaccharide chain to the protein. Finally, the transferred oligosaccharides are processed to their final structure. Different cellular compartments are involved in this pathway of glycoprotein biosynthesis. Synthesis and *en bloc* transfer of the precursor oligosaccharide are carried out in the endoplasmic reticulum (ER). Subsequent processing of glycan structure occurs in the Golgi apparatus.^[12]

Synthesis of dolichol-linked precursor oligosaccharide

The first step in *N*-linked glycosylation is the biosynthesis of the lipid linked oligosaccharide (LLO) using dolichol as a carrier for initial glycan structure. Dolichol consists of isoprenoid units delivering a lipid function which localize the biosynthetic pathway to the membrane of ER. Alterations in biosynthesis of dolichol can severely affect the

N-glycosylation process and further investigations have shown that the chain length varies among different species.^[13]

Biosynthesis of LLO is performed by a series of glycosyltransferases using nucleotide activated sugars (UDP-GlcNAc, GDP-Man, and UDP-Glc) as substrates, shown in Figure 3. All glycosyltransferases (GTf) involved in *N*-glycan precursor synthesis are encoded by *ALG* (standing for *altered* in *glycosylation*) genes. The process is initiated on the cytoplasmic face of the ER. Herein the anhydride dolichyl-pyrophosphate-GlcNAc (Dol-PP-GlcNAc) is produced by a transfer of GlcNAc-P from UDP-GlcNAc to membrane-bound dolichol-phosphate (Dol-P). The second GlcNAc residue is added by a protein complex encoded by the *ALG13* and the *ALG14* genes.^[14]

The synthesis process of LLO on the cytoplasmic side is finished by transferring five mannose residues from GDP-Man to the precursor. For this procedure three different enzymes are required to complete the $\text{Man}_5\text{GlcNAc}_2$ oligosaccharide structure. The transfer of the first mannose to the Dol-PP-GlcNAc₂ complex is initiated by β -1,4 mannosyltransferase (*ALG1* gene). The next two Man residues are subsequently attached in α -1,3 and α -1,6 position whereby this branching is catalysed by a single enzyme, encoded by *ALG2*. The last two mannoses become attached to the Man in α -1,3 core position using the glycosyltransferase encoded by *ALG11* gene. This proceeding is the last transfer accomplished on the cytoplasmic side of ER.^[15]

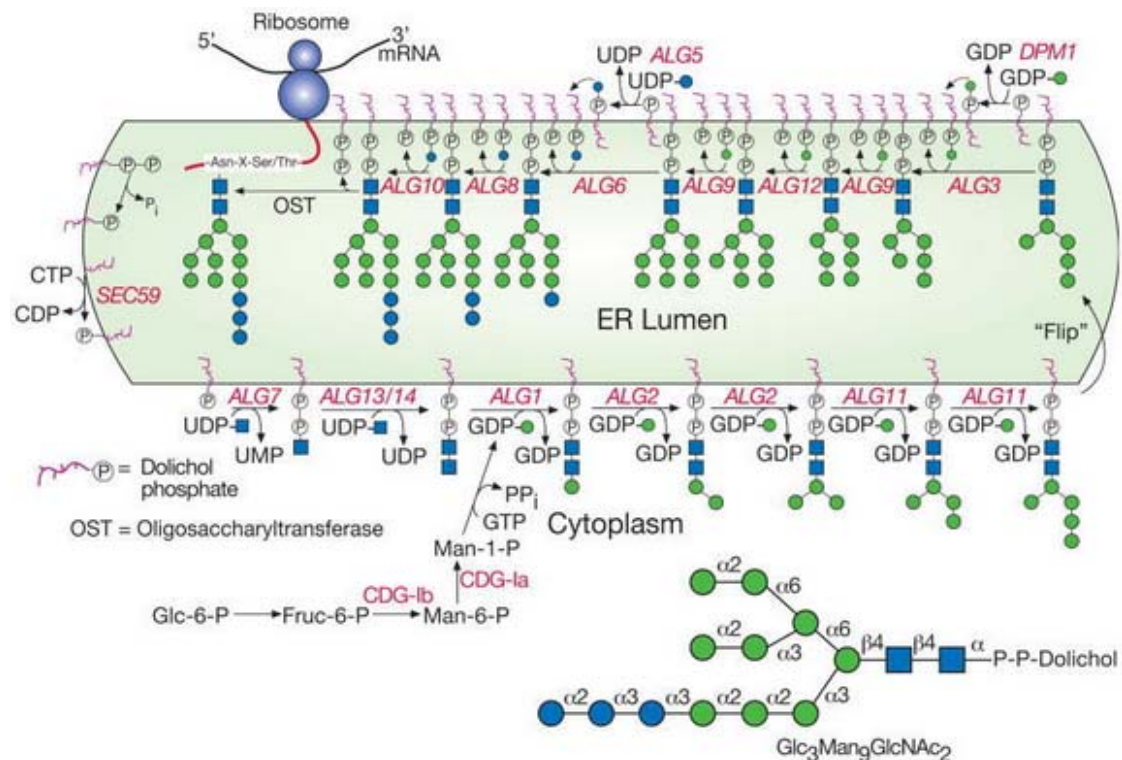


Figure 3: Synthesis pathway of $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2\text{-P-P-dolichol}$. The biosynthesis process starts on the cytoplasmic side of the ER by adding two GlcNAc and five Man to Dol-P. The $\text{Man}_5\text{GlcNAc}_2$ oligosaccharide is then flipped to the ER lumen where four Man and three Glc are added to complete the $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ precursor. All involved genes, substrates and localization of each single synthesis step are illustrated in the picture. Figure taken from *Essentials of Glycobiology*, Cold Spring Harbor, New York, 2009.^[16]

Subsequently a mechanism not fully understood yet translocates the $\text{Man}_5\text{GlcNAc}_2\text{-P-P-Dol}$ precursor across the membrane bilayer to the luminal side of the ER. Luminal biosynthesis adds four mannose and three glucose residues to the LLO using Dol-P-Man and Dol-P-Glc dependent glycosyltransferases. First b- and c-antenna are build up by adding mannose residues to complete the $\text{Man}_9\text{GlcNAc}_2$ precursor. This process is initiated by the α -1,3 mannosyltransferase encoded by the *ALG3* locus.^[17] If construction of $\text{Man}_9\text{GlcNAc}_2$ oligosaccharide has been finished glucosylation of 3-antenna is carried out by adding an α -1,3 linked glucose through *ALG6* glucosyltransferase.^[18] The addition of two further glucose molecules completes the biosynthesis of the mature N-glycan precursor $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2\text{-P-P-Dol}$.

Now the glycan is ready for attachment to proteins which have been translocated across the ER membrane. Deficiencies in LLO assembly may result in an accumulation of intermediates and hypoglycosylation of proteins leading to diseases observed in patients with Congenital Disorder of Glycosylation (CDG).^[19]

***En bloc* transfer of oligosaccharide to protein**

Oligosaccharyltransferase (OST) is the central enzyme for catalysing the transfer of glycan from lipid carrier to the amide group of an asparagine in the protein. OST is a protein complex which binds the membrane anchored Dol-P-P-glycan and transfers the oligosaccharide to the nascent protein by cleavage the high-energy GlcNAc-P bond. Thereby the Dol-P-P molecule is released and recycled by transporting it back to the cytoplasm.^[13]

The OST has been investigated most extensively in yeast (*Saccharomyces cerevisiae*). Protein purification processes have identified a heterooctameric yeast OST complex consisting of different membrane-bound subunits denoted by their gene names Ost1p, Ost2p, Ost3p or Ost6p, Ost4p, Ost5p, Wbp1p, Swp1p, and Stt3p. In mammals, three OST complexes have been identified which differ in their abilities to transfer the glycans.^[11] Protein sequence database searches indicate that complexes of both yeast and vertebrate have a very similar subunit composition. In mammals ribophorin I and II, OST48, and DAD1 have been identified which are related proteins to Ost1p, Swp1p, Wbp1p, and Ost2p, respectively. Wacker *et al.*^[20] has shown that Stt3p subunit contains the catalytic site of the OST complex responsible for *en bloc* N-glycosylation of asparagine. In mammalian cells two isoforms of the Stt3p protein exist showing a tissue-specific difference in relative expression. Such differences in isoform expression may indicate some regulatory effects.^[21]

Processing of oligosaccharide structure

After transfer, oligosaccharide precursor passes through a sequence of several assembly steps to reach the final, mature structure. Processing of glycan begins with the sequential removal of the three terminal glucose residues by glucosidases in ER (α -1,2 glucosidase I and α -1,3 glucosidase II). During this process, the innermost glucose residue can be recognized by calnexin and calreticulin, two lectins which vali-

date correct glycoprotein folding. If folding is accurate, the glycoprotein is allowed to leave the ER and enter the Golgi apparatus.^[22] The last reaction step in the ER is the removal of a mannose residue from the central arm catalysed by a specific α -1,2 mannosidase.^[23]

The further process right up to the mature glycoprotein involves *cis*-, *medial*-, and *trans*-Golgi. In the *cis*-Golgi compartment the oligosaccharides are trimmed again by mannosidases to a $\text{Man}_5\text{GlcNAc}_2$ intermediate playing a key role in assembly of hybrid and complex type *N*-glycans. Those structures which are not fully processed to $\text{Man}_5\text{GlcNAc}_2$ cannot undergo remodelling to form hybrid and complex structures.^[11]

Biosynthesis of hybrid and complex types is initiated in the *medial*-Golgi by adding an *N*-acetylglucosamine residue to the α 1-3 core mannose. This opens the possibility that both remaining Man on the core can be removed by α -mannosidase II. Once the α 1-6 mannose in the core is free, a second *N*-acetylglucosamine is added to the α 1-6 mannose. All further processes necessary to complete the complex *N*-glycan are fulfilled in *trans*-Golgi. Hybrid *N*-glycans are formed if the two terminal mannose residues are not removed by α -mannosidase II.^[11]

1.1.2 Biological Roles of Glycans

Generally speaking, biological functions of glycans can be divided into two major categories, *i.e.*, (1) structural and modulatory properties and (2) specific recognition. Most commonly, glycans get recognized by intrinsic and extrinsic glycan-binding proteins (GBPs). Intrinsic GBPs recognize glycans from the same organism, whereas recognition by extrinsic GBPs refers to different organism.^[24]

Carbohydrate structures are evolutionary stabilized by conferring beneficial physical and structural effects to proteins to which they are attached. For instance, *N*-glycans play an important role in correct protein folding. Furthermore, glycosylation has the ability to modify properties such as protein stability and solubility. It has been shown that often an increased carbohydrate content in proteins significantly enhance serum half-life and *in vivo* activity. Therefore, strategies to introduce *N*-linked glycosyla-

tion sites (glycoengineering) have been established in designing protein therapeutics.^[25] Further, glycosylation plays an important role in cell adhesion, migration, and cell-cell communication. Hence, most receptors on the cell surface possess *N*-glycosylation sites to act either as a positive or negative regulator in cell-cell interactions. For example, integrins modified by *N*-acetylglucosaminyltransferase III (GnT-III) inhibit cell migration and cancer metastasis whereas a modification catalyzed by GnT-V promotes cancer invasion.^[26]

Glycosylation plays a crucial role in the context of immune response. Pathogens carrying glycans mimicking those of their host cell surfaces, evade by this strategy the immune response of their hosts.^[24]

High diversity of glycosylation makes it possible that glycans are involved in diverse biological functions. For one given glycan different roles have been observed in different tissues and cell types, and changes in structure might occur during development and cell differentiation. Such alterations have been identified in certain diseases like cancer and inflammation giving glycosylation the potential role of biomarkers. This role has enormously stimulated the field of glycosylation research over the last decades.

1.1.3 Glycoproteomics in Disease Research

Over the last years, glycoproteomics has become a fast growing field in biomarker research. Novel applications using selective enrichment of glycoproteins combined with high-resolution MS have enabled sophisticated strategies in studying changes in glycosylation and glycan isoforms. Such investigations have shown that aberrant glycoforms detected in plasma and different tissues can be associated with various types of cancers and several neurological disorders.^[27]

Hua *et al.*^[28] developed an approach for profiling glycan structures of cell membranes to enable an alternate route toward cancer diagnosis. Therefore a chip-based PGC nano-LC-TOF/MS was used to quantitatively determine *N*-glycans from 15 different cancer cell lines. Based on identified glycans, changes in relative abundance of broad glycan classes (high mannose, complex/hybrid fucosylated, complex/hybrid sialylated, etc.) can be taken to differentiate between cell lines.

In MS-based cancer glycoproteomics lectins are widely used for selective enrichment or to target specific glycans as biomarker. For instance, *Aleuria aurantia* lectin (AAL) and *Sambucus nigra* agglutinin (SNA) have been applied in lectin affinity chromatography (LAC) workflow to enrich certain glycoproteins with fucose and sialic acid from different breast cancer cell lines. Statistical analysis has found at least 100 glycosites specific for the more aggressive (triple negative) tumor subtype.^[29]

Beside cancer, irregularities in glycosylation patterns have been observed in neurological disorders such as Alzheimer disease (AD) and Parkinson disease. Investigations^[30] of proteins from cerebrospinal fluid (CSF) in AD and in healthy individuals revealed the presence of less sialylated proteins in AD patients than in non-AD patients. Incorrect glycosylation of amyloid precursor protein (APP), a key player in developing AD, has been associated with AD too. γ -secretase and β -secretase, two enzymes required for generating amyloid β -peptide from APP also play an important role in protein glycosylation, as these enzymes affect the extent of *N*-glycosylation and sialylation of APP being therefore responsible for correct processing of the protein.^[31]

2. Enrichment Strategies for Proteins and Glycoproteins

The most challenging problem to overcome in clinical relevant biomarker research is the tremendous dynamic range of protein abundance in biological samples like human plasma or cell supernatant (SN). For instance, serum albumin at the high end and interleukin 6 at the low abundance end of human plasma differ by a factor of 10^{10} . Moreover, the top 22 most abundant proteins in plasma represent nearly 99% of the total protein mass, whereas proteins and glycoproteins of diagnostic interest exist in far lower abundance compared to remaining proteins.^[32] Therefore, enrichment strategies such as affinity chromatography (AFC) are necessary to overcome the problem of the high dynamic range of proteins and glycoproteins.

2.1 Principle of Affinity Chromatography

Affinity chromatography can be defined as a type of liquid chromatography where separating process of biochemical mixtures is based on highly specific and, under appropriate conditions, strong interactions between affinity ligands and analyte molecules. Thereby, a wide variety of binding agents such as antibodies, proteins, or DNA can be used as affinity ligands which are usually immobilized at the surface of the packing material filled into a column. The appropriate choice of the ligand, responsible for selectively binding of a given target or group of targets in the sample, is an important factor determining the success of the AFC method. With respect to the immobilized ligand, AFC can be divided into several categories such as LAC, boronate AFC, immunoaffinity chromatography, and immobilized metal ion affinity chromatography (IMAC).^[33]

Affinity-based chromatography can be carried out in several modes such as AFC with step elution, and affinity depletion. These implementations differ in their performance and on the targets.

2.1.1 Affinity Enrichment in Step Elution Mode

The step elution mode, also known as the on/off elution format, is the most common applied scheme in affinity chromatography separation. The whole process can be divided into four steps, namely binding, washing, elution, and regeneration (Figure 4). First, a sample mixture dissolved in a special application buffer is injected onto the affinity column. The pH and ionic strength of the application buffer mimic the native environment of the affinity ligand and its target. If appropriately chosen, only analyte molecules are retained by the column, while other compounds having no interaction to the ligand pass through the column unretained. After this process several washing steps with application buffer can be applied to get rid of all other components. During the next step an elution buffer is applied to dissociate the target molecules from the ligand. The elution buffer usually induces a change in pH or adds competing agents to displace analytes from the surface. Depending on the applied system, target molecules can be collected for later analysis or monitored directly by an on-line method using a high-performance liquid chromatography (HPLC) support.

Due to its simplicity, flexibility, and selectivity the on/off mode has been in wide use in analytical chemistry. This method easily allows carrying out selective purification and isolation of a target compound out of complex samples.^[34]

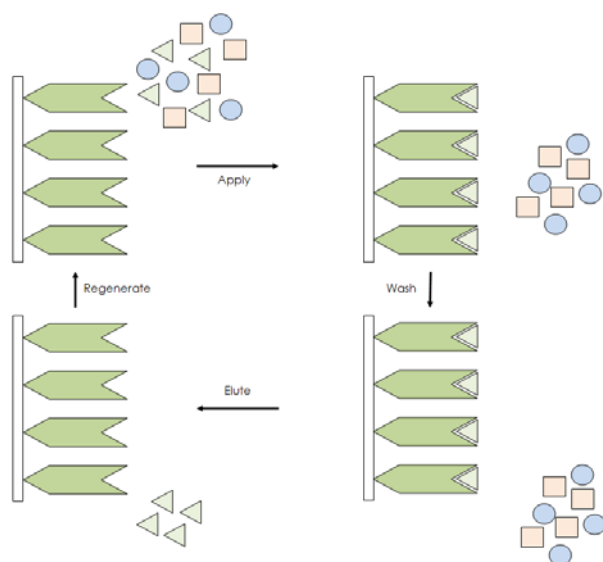


Figure 4: Step elution (on/off) format of affinity chromatography and the different steps involved. First target compounds get bound to the ligand whereas all other components are washed away. Upon applying the elution buffer the analyte molecules are eluted. Afterward the column is regenerated.

2.1.2 Affinity Depletion

In affinity depletion format a special affinity column with certain antibodies as ligands is used to get rid of e.g. the most abundant compounds from a complex sample. Subsequently, the non-retained sample components are analyzed by a second method.^[33] Affinity depletion methods are typically applied in proteomics issues to remove highly abundant proteins such as human serum albumin and IgG from serum to allow an analysis of lower abundant proteins in such samples.^[35, 36]

Complex biological samples exhibit a high dynamic range of proteins over several orders of magnitude. Thus, affinity depletion with antibodies against a few proteins is probably insufficient to solve this problem. A new method called combinatorial peptide ligand libraries (CPLL) was recently developed by Righetti and coworkers^[37] and has become a promising approach for low abundance protein analysis. In this method several millions of hexapeptides are used in an array of affinity ligands to enhance the relative and absolute abundance of low abundant species. The different 3-D structures of proteins usually show strong bio-affinity for just one hexapeptide and so the CPLL method allows investigation of at least eight orders of magnitude of complex samples.^[38]

2.2 Specific Enrichment of Glycoproteins and Glycopeptides

Different strategies are available for the specific enrichment of glycoproteins covering hydrazide capture, boronic acid AFC, and LAC. The hydrazide capture method is a solid-phase extraction for glycoprotein isolation, developed by Zhang *et al.*^[39] The principle behind this method is that cis-diol groups of carbohydrates get oxidized to aldehydes which then react with hydrazide groups forming covalent hydrazone bonds. For subsequent analysis, N-linked glycopeptides are treated with PNGase F to release the peptides which are determined by RP-HPLC-MS/MS afterward.

Boronic acid generate stable cyclic esters by with molecules containing cis-diol groups. This opens the possibility to employ boronic acid chemistry for specific glycopeptide enrichment.^[40]

LAC seems to be the most popular enrichment strategy for glycoproteins and glycopeptides. LAC coupled with MS based strategies has been widely performed in analytical assays for glycoprotein identification and glycopeptide analysis.^[41, 42, 43]

2.2.1 Lectin Affinity Chromatography (LAC)

Lectins are a diverse group of proteins showing characteristic carbohydrate-binding capacity with selectivity to certain sugar moieties. To date, there are at least 160 different lectins established from which more than 60 are commercially available to recognize diverse sugar structures. Table 1 lists some of the commonly used lectins for glycoprotein/glycopeptide enrichment together with their glycan-binding specificity.^[44] For example, Concanavalin A (Con A), the most extensively used lectin, recognizes the trimannosidic core structure of *N*-glycoproteins if these mannoses are not “blocked” by other residues. Con A also reacts with branched α -mannosidic structures of high-mannose and hybrid types.^[45] However, binding to tri- and tetra-antennary complex type glycans is low. Another well-known lectin is wheat germ agglutinin (WGA) which has an affinity for chitobiose *N*-acetylglucosamine and sialic acid. Thus, both lectins, Con A as well as WGA, have a broad specificity range what can be advantageous for exploring larger portions of the glycoproteome, whereas other lectins like SNA are more specific.

Lectin	Specificity
Concanavalin A (Con A)	High-mannose type, branched α -mannosidic structures
Wheat germ agglutinin (WGA)	<i>N</i> -acetylglucosamine; chitobiose (sialic acid)
Jacalin (JAC)	Galactosyl (β -1,3) <i>N</i> -acetylgalactosamine (<i>O</i> -glycoproteins)
<i>Sambucus nigra</i> lectin (SNA)	Sialic acid attached to terminal galactose in (α -2,6)
Peanut agglutinin (PNA)	Galactosyl (β -1,3) <i>N</i> -acetylgalactosamin (T-Antigen)
<i>Lens culinaris</i> agglutinin (LCA)	α -Linked mannose residues
<i>Phaseolus vulgaris</i> leucoagglutinin (PHA-L)	Tri/tetra-antennary complex-type <i>N</i> -glycan
<i>Aleuria aurantia</i> lectin (AAL)	Fucose linked (α -1,6) to <i>N</i> -acetylglucosamine; fucose linked (α -1,3) to <i>N</i> -acetylglucosamine

Table 1: List of some selected commonly used lectins and their glycan-binding specificities.

Table taken from Fanayan *et al.*^[44]

Binding conditions are dependent on the lectins applied. For example, Con A, binds more efficiently at a lower pH (~ 5.5), while many other lectins need neutral pH conditions. Also certain metals, such as calcium or manganese ions are required by lectins having specificity toward mannose structures. For desorption of bound glycoproteins from the AFC column, competitive elution is commonly performed using specific saccharide displacers (e.g. mannoside for Con A).^[44]

If analysis of the whole glycoproteome is desired, enrichment using only one single, selectively binding lectin is not enough. To overcome this drawback of selective binding, serial lectin affinity chromatography (SLAC) technique was developed for this purpose by Cummings and Kornfeld.^[46] Thereby, several lectin based affinity columns are used in tandem to retain different subsets of glycoproteins simultaneously. Multilectin affinity chromatography (MLAC) is a similar method to SLAC where the used column contains a physical mixture of different immobilized lectins.^[47] In comparison with SLAC, MLAC shows an enhanced binding affinity due to the combination of lectin binding sites and carbohydrate recognition units on surfaces, established as the so called "cluster glycoside effect".^[48]

3. Mass Spectrometry based Proteome Analysis

Mass spectrometry (MS) has evolved to an enormous and indispensable technique in analytical chemistry over the past century. An overview over this period shows that in the early 20th century, the technique was exclusively used by physicists to measure masses of atoms. Then, in the 1940s, mass spectrometers had become commercially available for industry to control production processes and in the 1980s, small organic molecules were routinely analyzed by MS though the technique was still far away from its state today.^[49] Only when soft ionization techniques such as matrix-assisted laser desorption/ionization (MALDI)^[50] and electrospray ionization (ESI)^[51] were developed MS became more and more attractive for analyzing macromolecules like proteins. Finally, the possibility of online-coupling between separation methods such as capillary electrophoresis (CE) or HPLC and tandem MS pushed these systems to the method of choice in high-throughput analysis of complex protein samples.^[52]

Several different types of MS instruments have been developed up to today. Thereby compromises between mass accuracy and mass resolution on the one hand, and the speed of data acquisition and sensitivity on the other hand, as well as the price had to be made. Within this development, the LTQ Orbitrap instrument, which is a hybrid type mass spectrometer consisting of an ion trap coupled to a Fourier transform (FT)-orbitrap analyzer, is one of the high-end instruments which enables MSⁿ capability with high resolving power, high mass accuracy and high sensitivity over a wide dynamic range. It allows accurate mass tandem MS (MS/MS) analysis of complex samples.^[53]

3.1 Electrospray Ionization (ESI)

ESI has become one of the most important ionization techniques in MS over the last three decades. This method was first described by Masamichi Yamashita and John Fenn in 1984.^[51] Thereby high voltage (usually 2-10 kV) is applied between a capillary tip filled with a liquid sample containing the analyte molecules and the entrance capillary to the mass spectrometer. In this way an aerosol is created and the ionized analytes are transferred into the gas phase. This offers the possibility for on-line cou-

pling liquid phase separations directly to a MS analyzer. The whole ionization process involves three steps starting with nebulization of a sample solution into electrically charged droplets. In the second step, the solvent is evaporated and, at the end, ions are ejected from highly charged droplets into the gas phase. These emitted ions are then accelerated into the mass analyzer.^[54]

3.1.1 Construction of ESI Source

Figure 5 displays a generally used construction of ESI in MS. Such systems consist of a needle positioned in front of the orifice of the mass spectrometer. For nebulizing the sample solution, a high voltage is applied between the top of the needle and the mass spectrometer. Dispersal of the solution results in a cone shaped (so-called Taylor cone) liquid emerging from the capillary end as first described by Geoffrey Taylor in 1964.^[55] ESI in its simplest way does not need more than the mentioned application. However, usually a coaxial nebulizing gas (N_2) flow is applied to assist building up the spray at higher flow rates. Compared with a pure electrospray, pneumatically assisted applications can handle higher flow rates at lower field strengths without the need of critical settings. The whole ionization process described in Figure 5 takes place under normal atmospheric pressure (API).^[56]

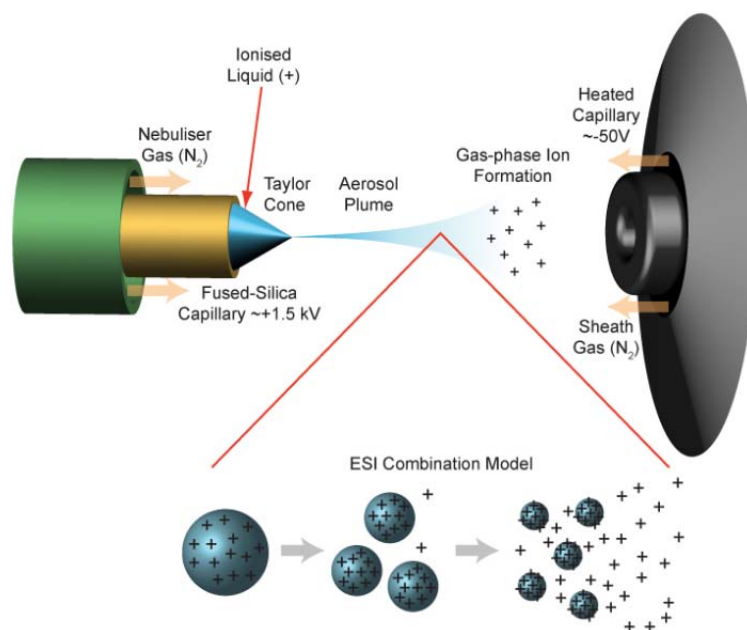


Figure 5: Scheme of an ESI source commonly used in MS. The sprayer needle contains the capillary delivering the sample solution as well as the nebulizing gas tube. It is positioned in front of the orifice of the mass analyzer. The Taylor cone which is built by ionised liquid is highlighted in blue. The figure also illustrates the ion evaporation model (IEM) resulting in multiply charged ions.

Picture taken from <http://www.lamondlab.com/MSResource/images/lcms/ESI.jpg> (03.11.2014).

3.1.2 Ionization Mechanism

A continuous stream of sample solution containing the analytes of interest is passed through a capillary to the tip of the needle kept at high voltage. At the pinpoint the liquid gets dispersed by electrospray resulting in highly charged droplets which exhibit the same polarity as the capillary voltage has. To minimize the initial droplet size, compounds which increase the conductivity are added to the solution (in positive mode for instance acetic acid or formic acid (FA)). The acidic character of these additives provides also a source of protons to facilitate the ionization process. Solutions used in negative mode should contain basic compounds like ammonium hydroxide.^[56] Solvents for ESI usually contain volatile organic compounds (e.g. acetonitrile (ACN) or methanol) to facilitate solvent evaporation that leads to a decreased droplet size and thus to an increased surface charge density. The evaporation process is continued as long as droplets reach the so called Rayleigh limit. At this point, the electrostatic repulsion of the ions becomes more powerful than the surface tension which holds the droplet together. As a consequence, the droplets undergo cou-

lomb explosion and, finally, ions at the droplet surface are ejected into the gaseous phase. The ion evaporation model (IEM) and the charge residue model (CRM) are the two major theories explaining the phenomenon behind the final production of gas-phase ions. The CRM theory originates from Dole *et al.*^[57] describing the process as cycles of evaporation and fission until droplet reach a radius of about one nanometer. Droplets with this size usually contain only one analyte ion. The IEM theory suggests an assistance of field desorption of solvated ions through a high field strength occurring at certain droplet radii.^[58]

Depending on the flow rate, the ionization process can be distinguished between classical electrospray and nanospray. The nano-electrospray operates at flow rates below 500 nL/min and has some main advantages compared with normal electrospray. Benefits of these low flow rates lie in improved ionization efficiency through a much higher charge-to-volume ratio than in conventional electrospray sources.^[59] Nano-spray is commonly operated without any nebulizer gas stream.

Since ESI is applied to liquid samples, the technique is well suited to be coupled subsequent to a HPLC separation. This set-up has become a most powerful and fast technology in analyzing complex biological samples.^[60, 61]

3.2 Orbitrap Technology

The Orbitrap mass analyzing technology was developed by Alexander Makarov in 2000.^[62] It enables a third type of dynamic ion trapping beside the linear and the segmented ring way. Thereby ions get electrostatically trapped around a spindle-shaped central electrode similar to orbits of planets in the Solar system.^[62] Frequency derived by harmonic ion oscillations along the axial direction are detected using image current and fast FT algorithms to determine the mass-to-charge (m/z) ratio of the ions.^[63] The Orbitrap analyzer provides high resolution (60,000), high-mass accuracy (<3-5 ppm), and a wide dynamic range. With the Orbitrap technology these properties are attained without the need of an expensive and strong magnetic field used in FT ion cyclotron resonance (FTICR) spectrometer.^[64]

3.2.1 Construction of Orbitrap Mass Analyzer

Figure 7 schematically illustrates the construction of the Orbitrap mass analyzer. The Orbitrap consists of an outer barrel-like electrode (maximal inner diameter: 30 mm) which is positioned coaxial to an inner spindle-like electrode (maximal outer diameter: 12 mm). In the LTQ Orbitrap XL instrument (Figure 6), a curved linear ion trap (C-trap) is used to trap the ions in a small cloud and inject them through a narrow ion channel into the Orbitrap analyzer. Electrodes are positioned on both ends of the Orbitrap to produce a potential barrier that ions cannot leave in axial direction. The outer electrode is separated into two parts using a ceramic ring.^[53] Comparing the whole construction with the size of a one euro coin illustrates that the Orbitrap is a small analyzer (Picture B).

30

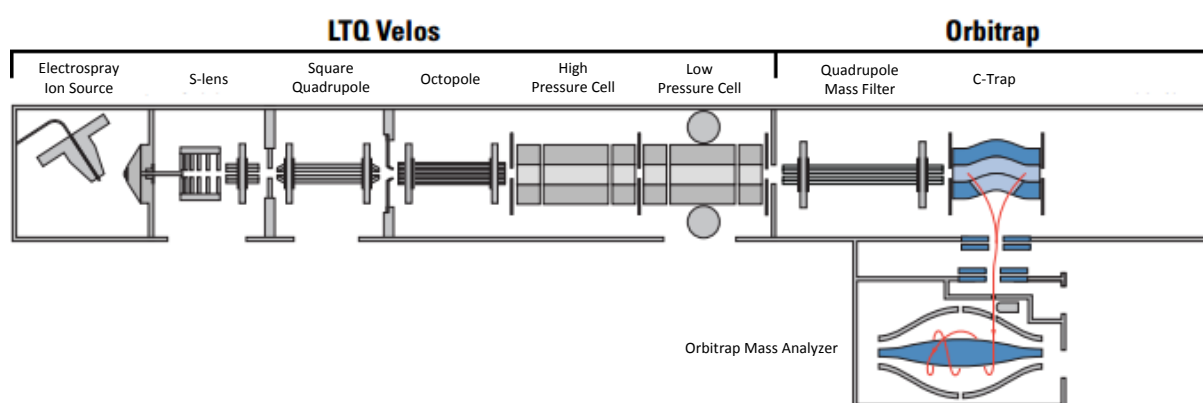


Figure 6: Construction scheme of the LTQ Orbitrap XL instrument without a HCD collision cell. After ionization, ions enter the linear trap quadrupole (LTQ) mass analyzer. After CID fragmentation processes carried out in the LTQ, the fragment ions can be analyzed under low resolution by the two detectors of the LTQ trap. If high resolution fragment ion analysis is required, the ions are focused and transferred into the C-trap. For accurate mass analysis, the ions are injected as a compressed ion cloud into the Orbitrap. Picture taken from Thermo Fisher Scientific (*LTQ Orbitrap Velos Hardware Manual*).

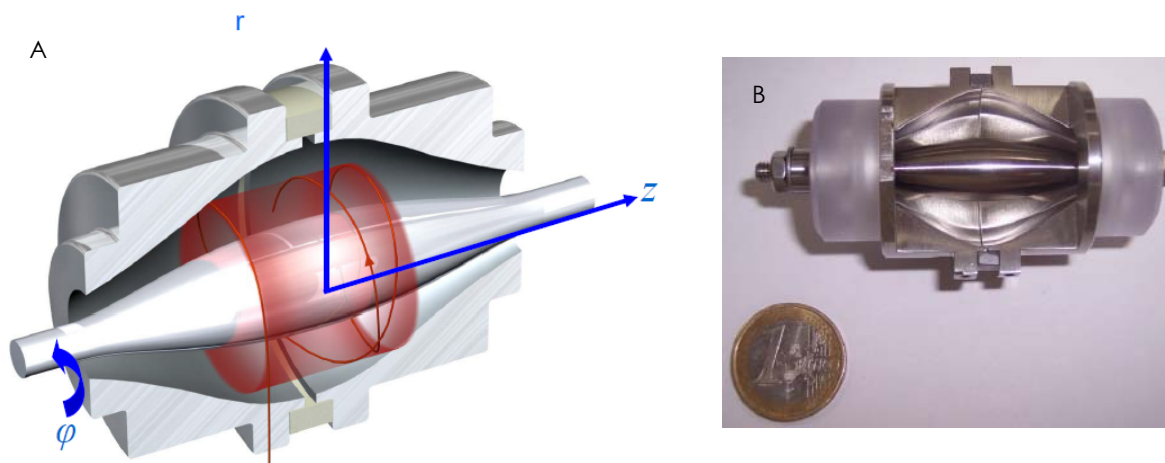


Figure 7: Schematic drawing of an Orbitrap mass analyzer. Picture A shows the trapping scheme of ions around the central spindle electrode. Picture B illustrates the size of the Orbitrap by comparison with a one Euro coin. Both pictures are taken from a presentation of Thermo Fisher Scientific Inc.

3.2.2 Theoretical Background

Before ions enter the Orbitrap they are cooled down in the C-trap by soft collisions with gas molecules *i.e.* ions lose their energy and become unable to escape. These collisions are mild enough to avoid any fragmentation. After this collisional cooling, ions form a thin, long thread which is compressed axially by applying 200 V to both outer-end electrodes. After that, DC pulses are applied to the electrodes to push out the ions orthogonally to the axis of the C-trap. Between the C-trap and the Orbitrap ions get accelerated and converged into a tight cloud by using appropriate ion optics. The ion clouds enter the Orbitrap tangentially through a small entrance aperture. After entering, the ions get trapped electrostatically around the central electrode by rapidly increasing the electric field. The raising field squeezes the trajectory closer to the axis. This process is continued until ions of all mass-to-charge ratios of interest have entered the Orbitrap. Then the voltage on the inner electrode is stabilized. At this point the ions rotate around the centre whereby lower m/z values are closer to the central electrode than higher ones.^[53, 64]

Additionally to the circle motion around the spindle, a harmonic oscillation of the ions in axial direction occurs as well. The frequency of this axial oscillation is characteristic for the m/z -ratio (cf. eqn. 1) of the ions and can be determined via the image

current induced by the axial oscillation of the ions. It is detected using a differential amplifier. The total image current registered delivers a transient derived from many ions of different m/z values present in the Orbitrap. The underlying frequencies building up this complex signal are determined by using FT. Out of the three characteristic frequencies ω_r (frequency of radial oscillation), ω_a (frequency of rotation), and ω_a (frequency of axial oscillation) only axial frequency is completely independent of ion energies and Orbitrap dimensions. Therefore, ω_a frequency is able to be used for determination of m/z values as the following formula illustrates:

$$\omega_a = \sqrt{\frac{k}{m/z}}$$

Eqn.1: ω_a = axial oscillation frequency; k = instrumental constant; m/z = mass-to-charge ratio ^[62]

3.3 Tandem Mass Spectrometry (MS/MS)

The idea behind the multistage MS strategy is that several mass spectrometric processes are carried out in tandem. First, an analyzer isolates certain precursor ions which are then fragmented into product ions and neutral fragments. After fragmentation, a second mass analyzer unit/process is used for analyzing the product ions. The essential step in tandem MS is the fragmentation of prior selected precursor ions. A number of fragmentation methods, some of them are discussed below, can be used to fragment the precursor ions. Different fragmentation methods give different types of fragmentation patterns and this information is accessible for structure identification.^[65]

3.3.1 CID Fragmentation

Today, collision-induced dissociation (CID) is the most common ion fragmentation technique in tandem MS. The CID fragmentation is understood as a two-step mechanism. First, a collision between a precursor ion and a neutral target gas (Ar, N₂) brings the ion into an excited state. This step is generally orders of magnitude faster

(10^{-14} to 10^{-16} s) than the second one. The higher internal energy of the ions leads in the following step to an unimolecular decomposition of the activated ion. Fragmentation of the precursor ion usually occurs if the collision energy is high enough that ion is excited beyond its threshold for dissociation.^[66] Under inelastic conditions only a fraction of the kinetic energy can be converted into vibrational energy. The following equation 2 describes this converted energy fraction.

$$E_{\text{con}} = E_{\text{kin}} \frac{M_t}{M_i + M_t}$$

Eqn.2: Maximum energy fraction converted into internal energy (E_{con}) in dependence on the kinetic energy of the ion in the laboratory frame of reference (E_{kin}) and the molecular masses of analyte ion (M_i) and the target collision gas (M_t).^[67]

The equation illustrates that the energy available for the vibrational activation will increase with the kinetic energy of the ion and the target mass of the collision gas. Furthermore, E_{con} decreases as a function of $1/M_i$. This means that larger precursor ions have less internal energy for collision induced fragmentation process available.^[67]

Depending on the collision energies, the CID process can be distinguished between low-energy collisions (eV range) and high-energy collisions (keV range). Low-energy CID is most often applied in ion traps or quadrupole collision cells (e.g. triple quad and QqTOF). In such instruments the collision chamber is a quadrupole filled with a neutral inert gas, usually N_2 or Ar, operated in radio frequency (RF) only mode for focusing of the fragment ion beam. Low energy CID of peptides mostly yields a cleavage of the amide bonds (besides some fragmentations of the other bonds of the peptide backbone). This amide-bond fragmentation leads to b- and y-ions depending on where the charge will be retained. Figure 8 illustrates that charged amino terms and carboxy terms are referred to b- and y-fragment ions, respectively. Therefore, low-energy CID is widely used in peptide sequencing and becomes an indispensable technique in proteomics issues.^[68]

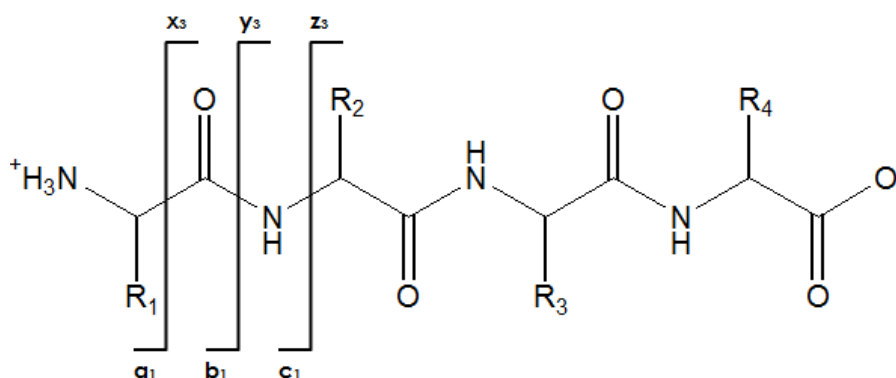


Figure 8: Nomenclature of fragment ions typically occurring in MS based peptide sequencing experiments. Applying low energy CID, b- and y-ions are predominantly obtained. ECD/ETD techniques lead exclusively to c- and z-ion fragments.

Higher energies for collision activation are usually applied in sectorfield or TOF-TOF (time-of-flight) instruments, where the precursor ions have a high kinetic energy in the keV range. Fragment spectra from peptides obtained by high-energy CID exhibit an increased side-chain fragmentation.^[69] Hence, low-energy techniques like surface-induced dissociation (eV SID) have also been implemented in TOF-TOF instruments to exploit advantages of both energy levels combined in a high resolution analyzer instrument.^[70]

With applications in ion traps the so called “one-third effect” occurs what is a system immanent drawback. It refers to a loss of fragment ions in MS/MS spectra if the fragment mass is less than one-third of the precursor ion m/z. Although CID remains the most common fragmentation type in MS, new activation methods like ECD (electron-capture dissociation), ETD (electron transfer dissociation), EDD (electron detachment dissociation) have been emerged in recent years, each with their own advantages and applications.

3.3.2 HCD Fragmentation

Higher-energy C-trap dissociation (HCD) is a specific CID fragmentation technique developed for a hybrid Orbitrap-MS instrument. In this instrument, a C-trap (curved linear ion trap) is used for capturing and focusing ions with a large mass range to inject them into the Orbitrap. In this construction, it was envisaged to use the C-trap additionally as a collision chamber to enable triple quadrupole-like fragmentation. For this purpose, a RF was raised up to 2,500 V (normal is 1,500 V) to retain as much fragment ions as possible. But experiments have shown that an increased RF leads to a decreased trapping efficiency of the C-trap. Therefore, Olsen *et al.*^[71] developed the LTQ Orbitrap XL instrument (Figure 9) where the HCD fragmentation are performed in an octopole collision cell (usually pressurized with nitrogen) at the far end of the C-trap. Since then, HCD has become an abbreviation for higher-energy collisional dissociation.

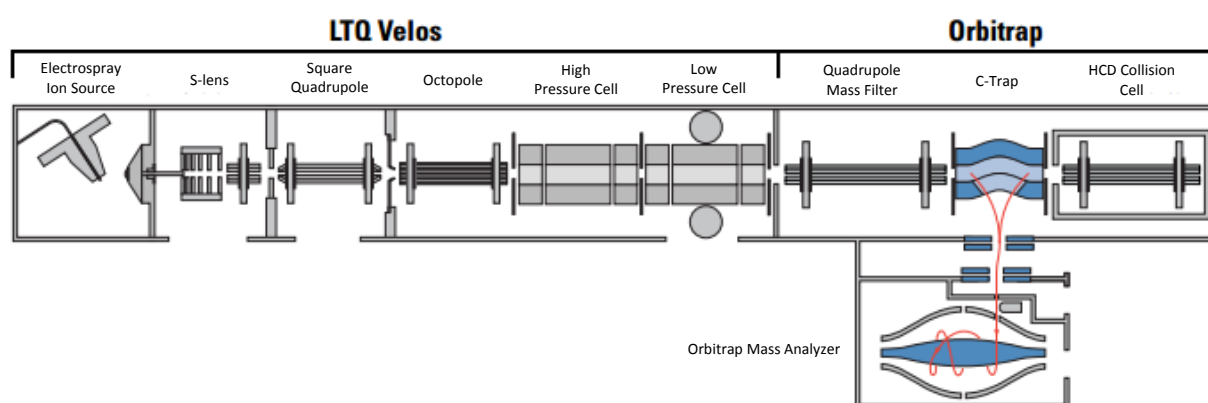


Figure 9: Construction scheme of the LTQ Orbitrap XL instrument. Alternatively to the precursor fragmentation in the LTQ under low energy CID conditions (described in Figure 6), the precursor ions can be transferred to the Octopole Collision Cell (HCD trap) where fragmentation under higher energy CID conditions takes place. In the latter case ion fragment analysis is carried out in the Orbitrap analyzer. Picture taken from Thermo Fisher Scientific (*LTQ Orbitrap Velos Hardware Manual*).

The HCD technology overcomes the one-third effect inherent to most ion traps and thus gets information about the low-mass region including characteristic b_1^- , b_2^- , y_1^- , and y_2^- -fragment ions becomes possible.^[71] Comparison of HCD and CID mode applied on the same peptides shows that HCD spectra contain smaller fragment ions and that fragmentation has a slight preference for producing y -ions with lower

charge states than in CID. Further differences occur most notably in the b-ion patterns obtained from doubly charged peptides. Occurrence of the singly charged b_2 -ion within the five most abundant peaks has a large probability in HCD spectra.^[72]

CID and HCD fragmentation applied on *N*- and *O*-linked oligosaccharide structures lead to two main fragmentation types, *i.e.* glycosidic bond cleavage and cross-ring cleavage. Cleavage of glycosidic bonds between two neighbouring sugar units yields B- and Y-fragment or C- and Z-fragment ions, respectively (Figure 10) and delivers information on the sequence and the branching of the structure. Cross-ring cleavage affects two linkages within the ring structure as indicated as superscripts. In positive ion mode, cleavages of the sugar ring occur rarely and need higher fragmentation energies like HCD, whereas glycosidic bond cleavage fragments are dominant in low-energy fragmentation types.^[73] In negative ion mode, cross ring fragmentation is common.

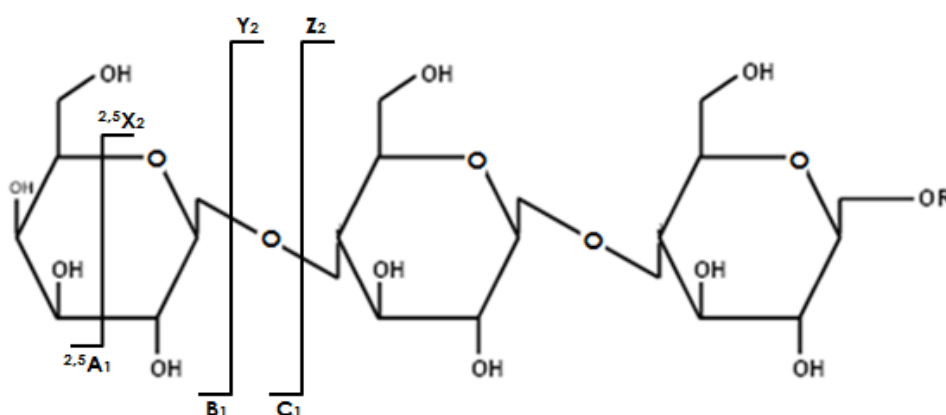


Figure 10: Nomenclature of fragment ions typical for MS based fragmentation of *N*- and *O*-linked oligosaccharides. Cross-ring cleavage occurs under higher fragmentation energies and delivers A- and X-fragment ions, affecting the two bonds indicated by the superscripts. Glycosidic bond cleavage results in B- and Y-ions or in C- and Z-ions, depending on which fragments carry the oxygen. Such fragments are observed more often under low-energy conditions.^[74]

3.3.3 ETD Fragmentation

The ETD process is an advancement of the ECD which was developed first by McLafferty and coworkers in 1998. In the ECD method, multiply protonated peptides or proteins are trapped in an ion trap or an ion cyclotron resonance (ICR) mass analyzer cell and are exposed to electrons possessing near-thermal energies. Capturing of such electrons causes a specific cleavage of the amine bonds (Figure 11) and leads to c- and z-product ions.^[75]

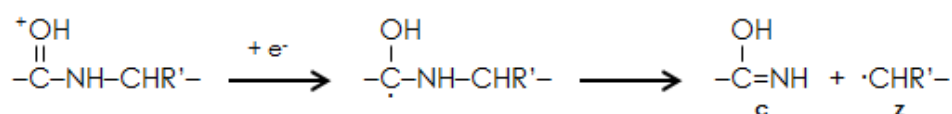


Figure 11: Mechanism of suggested radical site reaction. Captured electron leads first to a radical which are then fragmented in c- and z-ions. The process requires less energy than the b, y cleavage.

The ECD mode affects only the peptide backbone independently of the sequence. Posttranslational modifications (e.g. phosphorylation, sulfatation, glycosylation) are preserved by this method.^[76]

In MS instruments other than in ICRs, e.g. in ion traps where ions get caught in a RF electrostatic field, near-thermal electrons cannot be trapped too. Therefore, the ECD process requires FTICR instruments which come at a high financial cost. To overcome this problem, an electron transfer (ET) process was developed in which thermal electrons are first captured by an electron carrier molecule (EC process) which is subsequently moved into an ion-trap^[77] where the electron is transferred (ET process) to the analyte (acceptor) molecules (Figure 12). By these techniques the electron-capture/transfer process could be introduced to instruments more widely used in peptide and protein analysis such as linear and quadrupol ion traps, as well as quadrupol collision cells of QqTOF-instruments.

In an LTQ Orbitrap, the precursor ions are first trapped in the linear trap quadrupole and collected at the front section of the segmented trap. In the next step, the negatively charged carrier (or reagent) ions, usually fluoranthene radical anions, are injected from the end side of the linear trap and become mixed with the positive precursor ions. The electron transfer takes place within a time frame of 50 to 200 msec.

Essential for the ETD process is the production of the reagent radical anions which store the thermal electrons ready for transfer. Thermo Fisher Scientific solved this problem by using a filament producing fast electrons (> 70 eV) which are guided by magnets into the ion volume. The fluoranthene gas from a reagent vial enters together with an ultra-pure nitrogen carrier gas the ion volume where fast electrons collide with the nitrogen gas and produce positive nitrogen ions as well as slowed down thermal electrons (> 1 eV). Fluoranthene gas captures such electrons producing radical anions, which are then guided to the LTQ.

ETD fragmentation analysis of proteins carrying posttranslational modifications such as phosphorylation, sulfatation, O-, and N-linked glycosylation has shown similar results like the ECD fragmentation. Most notably ETD mode yields in c- and z-fragment ions. Hence, PTMs are usually not been affected by this fragmentation type.^[78]

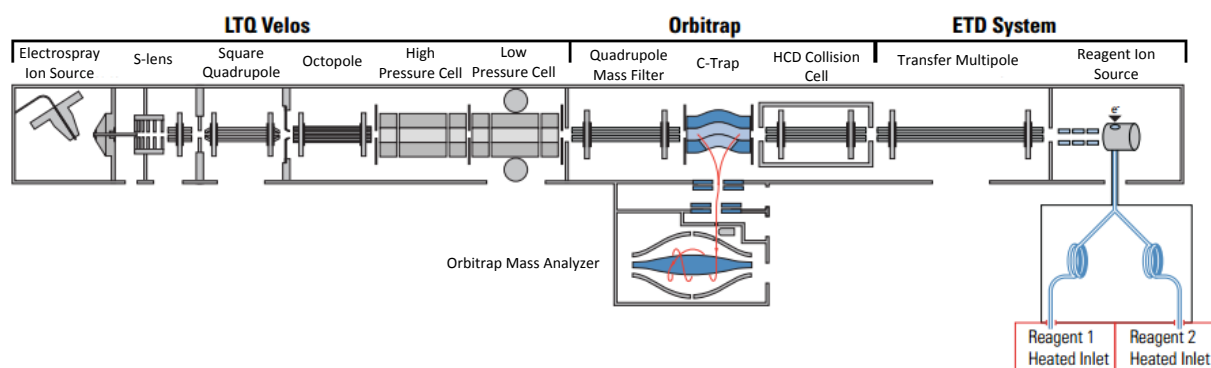


Figure 12: Construction scheme of the LTQ Orbitrap Velos instrument. Fluoranthene radical anions necessary for ETD fragmentation are produced in the reagent ion source. In this chamber fast electrons (> 70 eV) collide with a nitrogen carrier gas producing positive nitrogen ions and thermal electrons (> 1 eV). The slowed down thermal electrons are then captured by the fluoranthene gas to produce radical anions, which are then guided towards the LTQ, where ETD fragmentation takes place.

Picture taken from Thermo Fisher Scientific (*LTQ Orbitrap Velos Hardware Manual*).

V. II - EXPERIMENTAL WORK

1. Introduction

Glycosylation is one of the most common post-translational modifications in proteins. The glycoproteins can be differentiated in five different types whereby *N*- and *O*-linked glycosylation appear as the most frequent one. Thereby glycans are either attached to the amide group of asparagine (*N*-glycan) or to the hydroxyl group of serine or threonine (*O*-glycan). The attachment of the sugar chain to the asparagine in *N*-linked glycosylation needs an Asn-Xaa-Ser/Thr/Cys motif as consensus sequence where Xaa is any amino acid except proline.

The *N*-linked oligosaccharide structures can be classified into the complex, hybrid and high-mannose type (Figure 2). All types have in common the pentasaccharide core structure built up by two *N*-acetylglucosamine and three mannose molecules. High-mannose type only includes mannose molecules beside the core structure whereas the antennas in complex type contain different monosaccharide molecules. The hybrid type is a mixture of both complex and high-mannose type.

Glycoproteins can carry several different glycans on the same position of the protein. This means that amino acid sequence will be the same whereas the attached oligosaccharyl might be different. This is understood as the so called microheterogeneity. Due to the variety of possible glycans, glycoproteins are involved in many different biological processes such as cell-cell recognition, cell migration, cell adhesion, and signal transduction.^[79] Alterations in glycan structures of *N*-linked glycoproteins have been observed in various diseases, and therefore, changes in oligosaccharide structure of glycoproteins have become focused in biomarker research in recent years.^[80, 81] Such alterations might occur site-specifically.

In glycoproteomics, separation systems like HPLC coupled to MS/MS has become an indispensable technique for analyzing complex samples. The utilization of different fragmentation techniques such as CID, HCD and ETD on glycopeptides allows structure elucidation of both, glycan as well as peptide backbone.^[82]

CID fragmentation applied on *N*-linked glycopeptides in MS/MS techniques yields predominantly B- and Y-ions resulting from fragmentation of the glycosidic bonds within the glycan structure.^[74] To a minor extend, also C- and Z-ions are obtained. HCD fragment ions generated in the HCD-trap within the Orbitrap-Velos instrument are mass-analyzed in the Orbitrap-analyzer. Thus, molecular masses of oxonium ions originating from fragmentations of the glycan structures can be determined with a mass accuracy below 3-5 ppm. These monosaccharide, disaccharide and even trisaccharide oxonium ions are commonly used as indicator for the presence of glycopeptides.^[83] ETD fragmentation delivers only c- and z- fragment ions of the peptide backbone, whereas oligosaccharide structures and labile PTMs like phosphate groups are not affected by this technique.^[84]

In this Master thesis, we present a method for analyzing glycoproteins and alterations in their glycan structures present in complex protein samples obtained from the SN of MCF-7 (Michigan Cancer Foundation) cells. These cells were inflammatory activated with IL-1 β for monitoring eventually occurring alterations in glycan structures. SN of non-activated MCF-7 cells were used as control sample. To make the samples easier to handle, proteins were pre-fractionated to their molecular weight by SDS-PAGE. After cutting the gel into proper edges, the proteins were reduced, alkylated, and enzymatically digested with trypsin. Then the glycopeptides were analyzed by reversed phase liquid chromatography (RPLC) coupled to ESI-MS/MS using an LTQ Orbitrap Velos mass spectrometer and utilizing different fragmentation mechanisms like CID, HCD, and ETD.

One of the glycoproteins found up-regulated upon IL-1 β stimulation was thrombospondin-1 (TSP-1). This protein is a homotrimer glycoprotein in which each monomeric subunit consists of various domains. The protein belongs to a family of extracellular matrix (ECM) proteins and was first discovered in the early 1970s by Baenziger *et al.*^[85] Due to the domain structure and the different types of repeated domains (type 1, type 2 and type 3 repeats) (Figure 13) a wide range of cellular pathways and cell types get influenced by TSP-1 receptor binding.^[86] TSP-1 is regarded to play a role in certain diseases. For instance, protein levels are elevated in different cancer types and studies have shown that an overexpression in certain cell lines can decrease

tumor growth, whereas the presence of TSP-1 in the ECM has shown an increased tumor cell invasion.^[87]

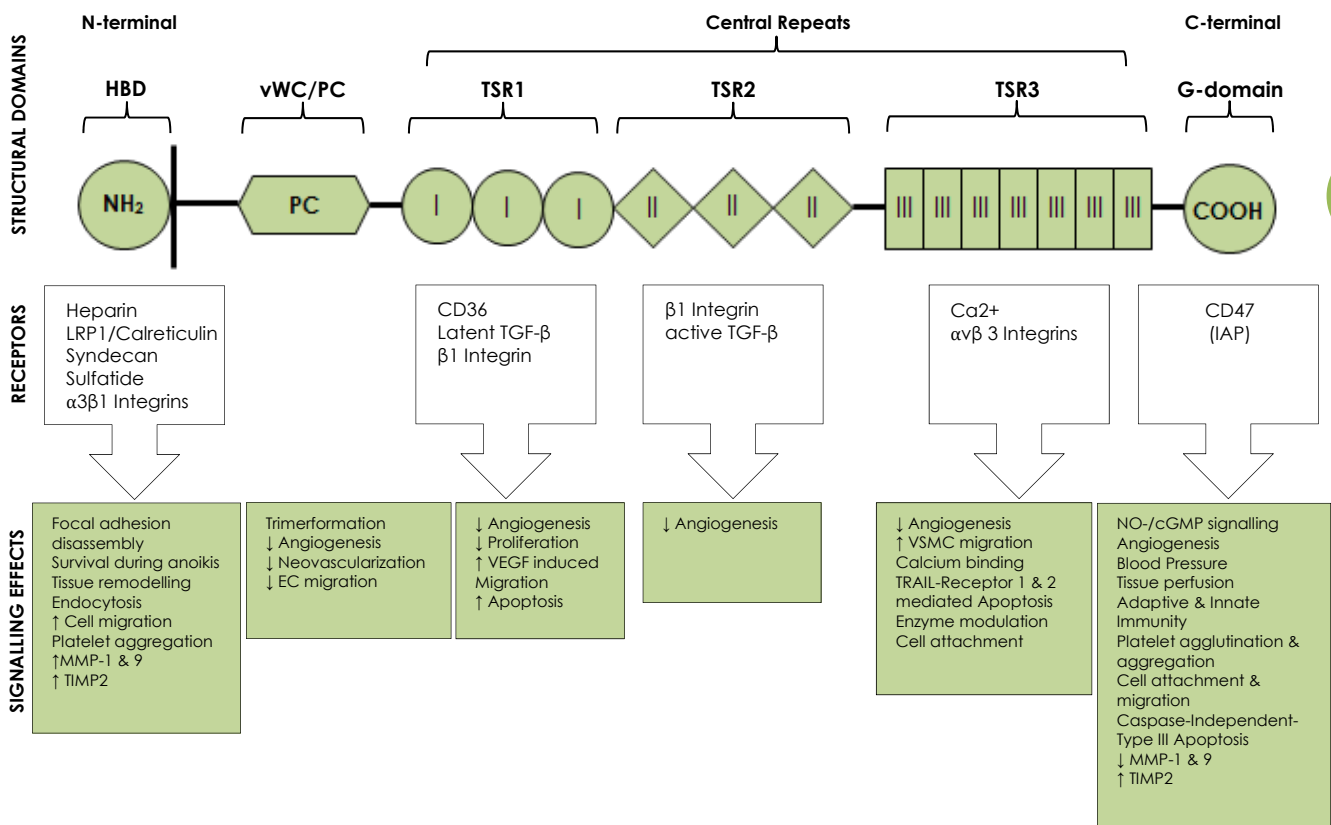


Figure 13: Structure of the thrombospondin-1 (TSP-1) monomeric subunit.

The figure shows the various domains and repeat types of TSP-1. Receptors affected by the different domains are listed below including potential functions thereof.

Figure taken from Krishna and Golledge.^[86]

2. Materials and Methods

2.1 Samples, Chemicals and Instruments

2.1.1 Samples

Cultured MCF-7 breast cancer cells were grown in Dulbecco modified Eagle's minimal essential medium (DMEM) supplemented with 10% fetal calf serum (FCS) and 100 U/ml penicillin/streptomycin. Cells were incubated at 37 °C and 5% CO₂ until confluence was reached. Inflammatory activation was carried out adding 10 ng/mL IL-1 β . Controls were cultured without adding IL-1 β .

2.1.2 Chemicals (Buffers and Solution)

- TEMED (N,N,N',N' – Tetramethylethylenediamine)
- APS (Ammonium persulphate)
- SDS (Sodium dodecyl sulphate)
- SeeBlue® Plus 2 Pre-stained Protein Standard
- DTT (Dithiothreitol)
- IAA (Iodoacetamide)
- Solvent A: Water with 0.1% FA
- Solvent B: 80% ACN, 20% Water and 0.08% FA

- **30% Polyacrylamide Solution**

292 g Acrylamide

8 g PDA (Piperazine di-acrylamide)

Ad 1 L ddH₂O

Store at 4 °C

- **2 M TRIS-HCl pH 8.8**

242.28 g TRIS

Ad pH 8.8 adjusted with HCl

Ad 1 L ddH₂O

Store at 4 °C

- **1 M TRIS HCl pH 6.8**

60.57 g TRIS

Ad pH 6.8 adjusted with HCl

Ad 1 L ddH₂O

Store at 4 °C

- **12% Separating Gel**

4.8 mL 30% Polyacrylamide Solution (ad 12%)

2.25 mL 2 M TRIS-HCl pH 8.8 (ad 375 mM)

4.83 mL ddH₂O

Immediately before usage:

50 µL 20% SDS Solution (ad 0.1%)

45 µL 10% APS Solution (ad 0.05%)

7.5 µL TEMED (ad 0.1%)

- **4% Stacking Gel**

1.06 mL 30% Polyacrylamide Solution (ad 4%)

1 mL 1 M TRIS-HCl pH 6.8 (ad 125 mM)

5.85 mL ddH₂O

Immediately before usage:

40 µL 20% SDS Solution (ad 0.1%)

40 µL 10% APS Solution (ad 0.05%)

8 µL TEMED (ad 0.1%)

- **5x SDS Sample Buffer**

5 mL 1 M TRIS-HCl pH 6.8 (ad 250 mM)

2 g SDS (ad 10%)

10 mL Glycerin (ad 50%)

0.05 g Bromphenol blue (ad 0.25%)

Ad 20 mL ddH₂O

Per 0.5 mL 5x SDS Sample Buffer: 70 µL β-mercaptoethanol (ad 12.5%)

- **Sample Buffer**

22.5 g Urea (ad 7.5 M)

5.7 g Thiourea (ad 1.5 M)

2 g CHAPS (ad 4%)

125 µL 20% SDS (ad 0.05%)

5 mL 1 M DTT (ad 100 mM)

Ad 50 mL ddH₂O

Store at -20 °C in proper amounts

- **10x TRIS-Glycine Buffer**

60 g TRIS (ad 25 mM)

288 g Glycine (ad 192 mM)

Ad 2 L ddH₂O

Store at RT

- **Electrophoresis Buffer**

100 mL 10x TRIS-Glycine Buffer

5 mL 20% SDS (ad 0.1%)

Ad 1 L ddH₂O

Store at RT

- **Fixing Solution**

500 mL Methanol (ad 50%)

100 mL Acetic acid (ad 10%)

400 mL ddH₂O

Store at RT

- **2% Sodium thiosulfate pentahydrate**

2 g Na₂S₂O₃ x 5 H₂O

Ad 100 g ddH₂O

Store at 4 °C

Before usage: 1:100 dilutions in ddH₂O

- **0.1% Silver nitrate Solution**

0.1 g AgNO₃

Ad 100 mL ddH₂O

Store at ice

- **Developer Solution**

3 g Na₂CO₃ (ad 3%)

130 µL 37% Formaldehyde (ad 0.05%)

Ad 100 mL ddH₂O

- **Wash Solution**

50 mL Methanol (ad 50%)

40 mL ddH₂O

10 mL Acetic acid (ad 10%)

Store at 4 °C

- **Destaining Solution**

1 mL 150 mM $\text{K}_3\text{Fe}(\text{CN})_6$ (ad 15 mM)
1 mL 500 mM $\text{Na}_2\text{S}_2\text{O}_3$ (ad 50 mM)
8 mL ddH₂O

- **Ammonium bicarbonate Buffer**

0.198 g NH_4HCO_3 (ad 50 mM)
Ad 50 mL ddH₂O

- **DTT Solution**

50 μL 1 M DTT (ad 10 mM)
5 mL 50 mM Ammonium bicarbonate buffer

- **IAA Solution**

500 μL 500 mM IAA (ad 50 mM)
4.5 mL 50 mM Ammonium bicarbonate buffer

- **Trypsin Solution**

20 μg Trypsin (ad 125 ng/ μL)
160 μL 1 mM HCl
Store at -20 °C
Before usage: 1:10 dilutions in ammonium bicarbonate buffer

- **Elution Solution**

50 mL ACN (ad 50%)
5.5 mL 90% FA (ad 5%)
Ad 100 mL ddH₂O
Store at 4 °C

2.1.3 Instruments

Instrument	Company
UltiMate 3,000 Nano LC Systems	Thermo Fisher Scientific Inc. Sunnyvale, USA
LTQ Orbitrap Velos	Thermo Fisher Scientific Inc. Waltham, USA
Acclaim® PepMap100 C18 column	Thermo Fisher Scientific Inc. Sunnyvale, USA
BioRad Mini-PROTEAN® Tetra System	Bio-Rad Laboratories Hercules, USA
Electrophoresis Power Supply EPS 301	GE Healthcare Life Sciences Buckinghamshire, GB
UNIVAPO 150 H Vacuum Concentrator	UniEquip Laborgerätebau- und Vertriebs GmbH Planegg, D
Sonorex	BANDELIN electronic GmbH & Co. KG Berlin, D
W6 Water Bath	Grant Instruments Cambridge, GB

Table 2: List of instruments used within the described experiments, and suppliers.

2.2 Methods

2.2.1 Cell Culture

MCF-7 cells were cultured in DMEM (Gibco, Life Technologies) complemented with 10% FCS and 100 U/ml penicillin/streptomycin (both ATCC, USA) and incubated at 37 °C and 5% CO₂ until confluence. Inflammatory activation was carried out adding 10 ng/mL IL-1 β (Sigma-Aldrich, USA) for 24 hours. After incubation, cells were washed once with 1xPBS buffer and cultured in 6 mL serum-free medium (RPMI 1640, Gibco, Life Technologies) for 6 hours. Subsequently, cell supernatants were filtrated using a 0.2 μ m filter (Whatman, Germany) and precipitated with the four-fold volume of ice-cold ethanol (Australco, Austria) overnight. MCF-7 cells were kindly prepared and the SNs provided by Andrea Bileck (Department of Analytical Chemistry, University of Vienna).

The SN-samples of the MCF-7 cells, either after treatment with IL-1 β or without treatment, were centrifuged for 25 minutes at 5,000 rpm and 4 °C. Subsequently, the pellet was completely dried by a desiccator for 10 minutes. Thereafter, the pellet was dissolved in sample buffer with additional urea (up to saturation) to facilitate the dissolving process. The protein concentration was determined by a Bradford assay.

2.2.2 SDS-PAGE for subsequent Shotgun Analysis

The different samples (IL-1 β treated and non-treated) were loaded onto a 12% polyacrylamide gel to separate proteins according to their molecular weight (Figure 14). A volume corresponding to 20 μ g of protein was loaded to each lane. The electrophoresis process was performed for 50 minutes at RT by a voltage of 250 V and a current of 20 mA until the prestained molecular markers (SeeBlue® Plus 2 Pre-stained Protein Standard [Invitrogen®, Carlsbad, USA]) were completely separated.

After electrophoresis, the gel was fixed with fixing solution for 30 minutes and subsequently a silver staining procedure was performed (described in 2.3.3)

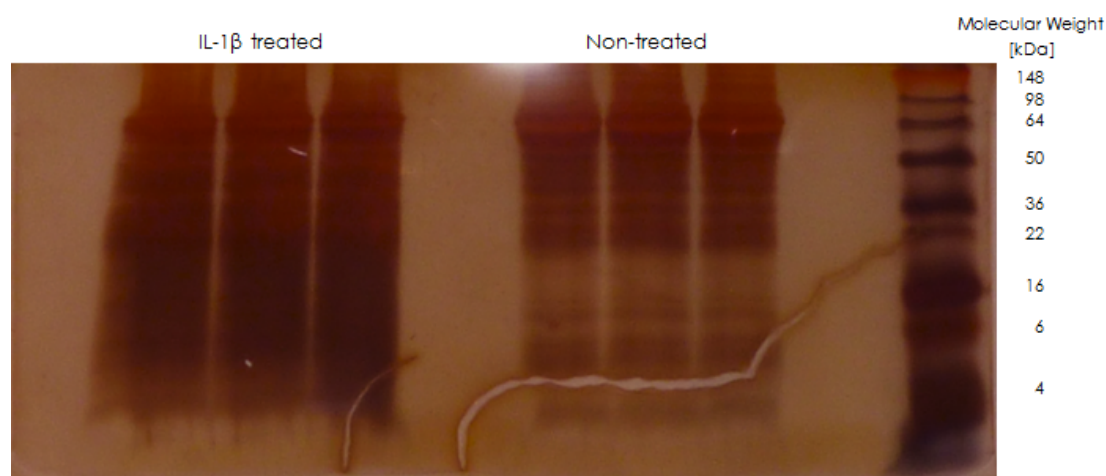


Figure 14: SDS-PAGE of SN samples of MCF-7 cells which have been treated by IL-1 β , or were non-treated. 20 μ g of proteins were applied to each lane. Proteins were separated according to their molecular weight using a 12% separation gel. A SeeBlue® Plus 2 Pre-stained Protein Standard (right-hand side) was used as marker.

2.2.3 Silver staining

For silver staining the gel was fixed with 50% methanol for 10 minutes followed by two washing steps using ddH₂O for 5 minutes each. Afterwards, the gel was sensitized by treatment with a 0.02% sodium thiosulfate solution. The sensitization was followed by washing steps using ddH₂O. Afterwards, the gel was stained applying a 0.1% ice cold silver nitrate solution for 10 minutes followed by two washing steps with ddH₂O. Subsequently, the gel was treated with the developer solution until precipitation was visible. Then, the reaction was stopped by adding 1% acetic acid.

2.2.4 Reduction, Alkylation, and Trypsin Digest

After silver staining, the gel area between 150 kDa and 40 kDa was cut into four equal edges as shown in Figure 15. Subsequently, these edges were chopped into proper pieces and collected in Eppendorf tubes. Then, pieces were destained by applying 300 µL destaining solution. The Eppendorf tubes were vortexed for about 5 minutes until pieces were completely destained. Afterwards, the destaining solution was removed and the gel pieces were washed with 200 µL of wash solution by shaking the tubes at RT and 800 rpm for 5 minutes. This step was overall carried out four times, continued by adding 200 µL ammonium bicarbonate buffer. Thereafter, disulphide bonds were first reduced with 200 µL 10 mM DTT (Sigma-Aldrich, USA) and then alkylated with 200 µL 50 mM IAA (Sigma-Aldrich, USA). The reduction process was carried out at 56 °C for 30 minutes. The IAA-alkylation reaction was incubated at RT for 20 minutes. In between and afterwards, the gel pieces were treated with 200 µL ammonium bicarbonate buffer by shaking at RT and 800 rpm for 5 minutes.

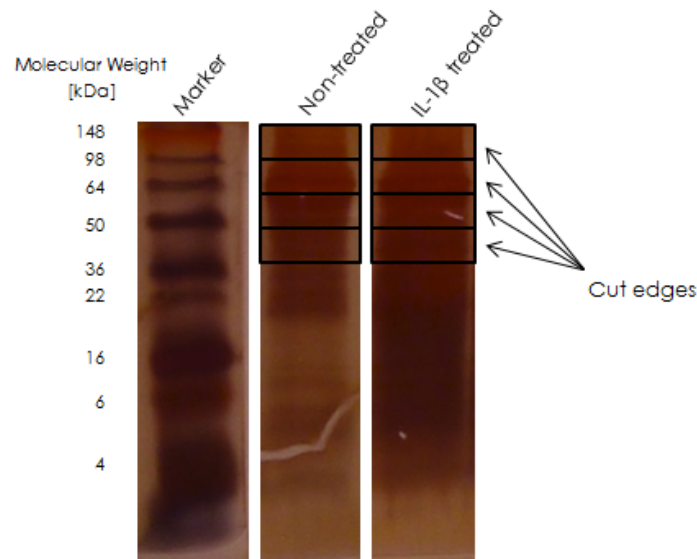


Figure 15: Gel sections containing non-activated and IL-1 β activated samples. The area between 150 kDa and 40 kDa was cut into four equal edges for subsequent analysis. Gel pieces were destained and proteins were reduced with DTT and alkylated by IAA. In-gel digest was carried out at 37 °C using trypsin solution for 17 hours.

Proteins were digested by adding 15 μ L of a 12.5 ng/ μ L trypsin solution (giving a 1:20 relation between trypsin and protein). Self-digestion of the enzyme was avoided by incubating the proteins first on ice for 15 minutes. In this time the gel pieces will soak up the solution without starting digestion. Thereafter, 25 μ L of ammonium bicarbonate buffer was added to the soaked gel. After that, the proteins were incubated at 37 °C over night for about 18 hours.

After digestion, 40 μ L ammonium bicarbonate buffer was added and then the gel pieces were sonicated for 15 minutes. The SN was collected in another Eppendorf tube. Thereafter, the gel pieces were treated twice by adding 40 μ L of elution solution and sonicated for 10 minutes. SNs were collected in corresponding Eppendorf tubes. At the end the eluted peptides were concentrated in the vacuum concentrator at 35 °C.

2.2.5 MS Analysis

For RP-HPLC-ESI-MS/MS measurement, lyophilised peptides and glycopeptides were dissolved in 25 μ L purified water (MS-grade) containing 2% ACN and 0.1% FA. Peptide mixture was analyzed by nanospray HPLC-MS/MS using an Ultimate 3,000 Nano LC System (Thermo Fisher Scientific) coupled to the LTQ Orbitrap Velos (Thermo Fisher Scientific) instrument. The sample was loaded onto an Acclaim® PepMap100 C18 column (75 μ m i.d. x 15 cm length) (Thermo Fisher Scientific) equipped with a pre-column of the same packing material (dimensions: 100 μ m i.d. x 2 cm length) and separated at a flow rate of 300 nL/min using a linear gradient of 7% to 35% solvent B (80% ACN and 20% water with 0.08% FA) in 30 minutes, followed by an increase to 40% solvent B in 2 minutes (held for 2 minutes), and then to 80% B in 2 minutes (held for 4 minutes). Solvent A was water containing 0.1% FA. The CID, HCD and ETD fragmentation was carried out, alternatively as specified below.

2.2.5.1 Top6 Method with CID Fragmentation

A Top6 method with CID fragmentation in the LTQ was developed by which the precursor ions giving the six most intense peaks were fragmented. With this method, a full scan of the MS¹ is acquired in the Orbitrap analyzer, whereas peptides and glycopeptides representing the six most abundant precursor ions are CID fragmented and analyzed in the LTQ. Further settings of this strategy are shown in Table 3. CID fragmentation applied to peptides is expected to deliver b- and y-fragment ions, whereas glycopeptides are expected to deliver B- and Y-ions.

MS¹	
Mass analysis	Orbitrap
Acquisition time	50 minutes
Scan range	400-1,400 m/z
Resolution	60,000
AGC target	1 x 10 ⁶
MS²	
Mass analysis	LTQ
Fragmentation type	CID (in LTQ)
AGC target	1 x 10 ⁴
Isolation mode	Quadrupole
Number of most intense peaks	6
Included charge state	2-5
Isolation window	3 m/z
Collision energy	35 eV
Activation time	30 ms
Dynamic exclude after n times	1 (for 60 seconds)

Table 3: Settings of the Top6 method for CID fragmentation.

Ions corresponding to the six most intense peaks within the scan range m/z 400 to 1,400 (acquired by the Orbitrap analyzer) are taken as precursor ions for fragmentation by CID within the LTQ using collision energy of 35 eV and an activation time of 30ms. After fragmentation, precursor mass will be written on an exclusion list for 60 seconds.

2.2.5.2 Data Dependent Top6 Method with HCD and ETD fragmentation

This method combines HCD and ETD fragmentation. The ETD fragmentation is triggered data dependent on the HCD spectrum. In this method, the precursor ions giving the six most abundant peaks in the full scan mode were fragmented in the HCD collision cell. The fragments are analyzed in the Orbitrap analyzer. If this fragmentation yields in fragments assigned as HexNAc (*N*-Acetylhexoseamine) and Hex-HexNAc molecules, an additional ETD event is applied on the same precursor, to get additional information about the peptide backbone of the glycopeptide. All settings of the method are outlined in Table 4.

MS¹	
Mass analysis	Orbitrap
Acquisition time	50 minutes
Scan range	400-1,400 m/z
Resolution	15,000
AGC target	1 x 10 ⁶
MS²	
Mass analysis	Orbitrap
Fragmentation type	HCD (in HCD trap)
AGC target	1 x 10 ⁴
Precursor selection range filter	750-1,400 m/z
Number of most intense peaks	6
Included charge state	3-5
Isolation window	3 m/z
Collision energy	27 eV
Activation time	10 ms
Dynamic exclude after n times	1 (for 60 seconds)
ETD	
Mass analysis	LTQ
Fragmentation type	ETD (in LTQ)
Activation time	75 ms + supplemental activation
Product masses selected for triggering ETD-process	m/z 204.087 (HexNAc) m/z 366.138 (Hex-HexNAc)

Table 4: Settings of data dependent Top6 method with HCD and ETD fragmentation.

Precursors of the six most intense peaks within the selection range m/z 750 to 1,400 are fragmented in the HCD trap using collision energy of 27 eV and an activation time of 10 ms. An ETD event (activation time 75ms) is only triggered if HCD spectrum contains the two B-fragment peaks m/z 204.087 (HexNAc) and m/z 366.138 (Hex-HexNAc).

The raw data file from the Top6 method with CID fragmentation was converted to a Mascot generic format (MGF) using the *ProteoWizard* (version 3.0.4778, 32-bit)^[88] software. This MGF file was used for identifying proteins which were present in the SN.

For this purpose the SearchGUI (version 1.19.5)^[89] of the *PeptideShaker* software (version 0.31.5) was employed applying the settings of Table 5.

<i>PeptideShaker</i> settings	
Fixed modifications	Carbamidomethyl C (Cysteine)
Variable modifications	Oxidation of M (Methionine)
Enzyme	Trypsin
Precursor mass tolerance	5 ppm
Fragment ion types	b- and y-ions
Max. missed cleavages	2
Fragment mass tolerance	0.5 Da
Precursor charge	2-4

Table 5: Settings for the *PeptideShaker* software used for protein identification:

The algorithm compares entries in a protein database with fragment ions in spectra by applying settings of the table.

3. Results and Discussion

3.1 Glycoprotein Identification in SN of MCF-7 Cells

Proteins present in the SN of MCF-7 cells were identified following the commonly used bottom-up approach. For this purpose the *PeptideShaker* software was used with settings shown in Table 5. The number of proteins and glycoproteins identified in control and IL-1 β activated samples, respectively, are listed in Table 6. Only proteins which have reached the highest calculated confidence level (100) were counted. The protein database UniProtKB was used to determine whether the recognized proteins possess an *N*-glycosylation site or not. Examination resulted in 72 and 88 *N*-glycoproteins in control and IL-1 β treated sample, respectively.

Number of Identified Proteins/Glycoproteins		
	Control	IL-1 β activated
Number of Proteins	248	258
Number of <i>N</i> -Glycoproteins	72	88

Table 6: Number of identified proteins and *N*-glycoproteins. The table lists the number of identified proteins and glycoproteins in non-activated and activated samples. Only proteins and glycoproteins which have reached the highest confidence level of 100 are listed.

Matching all *N*-linked glycoproteins identified in control and IL-1 β group showed that at least 47 of them occur in both groups. Table 7 shows an overview of all these glycoproteins carrying *N*-linked oligosaccharides together with their UniProtKB number.

Identified *N*-linked Glycoproteins

P07996	Thrombospondin-1	P02647	Apolipoprotein A-I
P01023	Alpha-2-macroglobulin	P05067	Amyloid beta A4 protein
P02768	Serum albumin	P05543	Thyroxine-binding globulin
P20742	Pregnancy zone protein	O00391	Sulfhydryl oxidase 1
O94985	Calsyntenin-1	O14786	Neuropilin-1
P01024	Complement C3	P00734	Prothrombin
P02751	Fibronectin	P00747	Plasminogen
P10586	Receptor-type tyrosine-protein phosphatase F	P55290	Cadherin-13
P19827	Inter-alpha-trypsin inhibitor heavy chain H1	Q04756	Hepatocyte growth factor activator
P19823	Inter-alpha-trypsin inhibitor heavy chain H2	Q08380	Galectin-3-binding protein
Q06033	Inter-alpha-trypsin inhibitor heavy chain H3	Q6EMK4	Vasorin
Q14624	Inter-alpha-trypsin inhibitor heavy chain H4	Q99715	Collagen alpha-1
P02765	Alpha-2-HS-glycoprotein	P08697	Alpha-2-antiplasmin
P49747	Cartilage oligomeric matrix protein	P04004	Vitronectin
P02771	Alpha-fetoprotein	P02774	Vitamin D-binding protein
P0C0L4	Complement C4-A	P07339	Cathepsin D
Q15262	Receptor-type tyrosine-protein phosphatase kappa	P36955	Pigment epithelium-derived factor
P01008	Antithrombin-III	P05156	Complement factor I
P13591	Neural cell adhesion molecule 1	O75144	ICOS ligand
Q13740	CD166 antigen	P00742	Coagulation factor X
Q92626	Peroxidasin homolog	P25311	Zinc-alpha-2-glycoprotein
P11047	Laminin subunit gamma-1	P10909	Clusterin
P02788	Lactotransferrin	P52823	Stanniocalcin-1
P32004	Neural cell adhesion molecule L1		

Table 7: Identified glycoproteins with *N*-linked glycosylation.

The table lists all glycoproteins with *N*-linked oligosaccharides which have reached the highest confidence level (100) by using *PeptideShaker* Search GUI.

For the identification of glycan structures, the MS² spectra after CID fragmentation were taken for a manually evaluation using *GlycoMod* software.^[90] First, spectra of glycopeptides were selected by searching for specific B-fragment ions typical for glycan fragmentation, (e.g. HexNAc-Hex (m/z 366.190) fragment). Based on these

identified spectra, the monoisotopic precursor mass $[M+H]^+$ was calculated and taken for *GlycoMod* searching algorithm. Monoisotopic precursor masses together with UniProtKB numbers of N-linked glycoproteins (Table 7) was considered to find possible glycopeptides applying settings of Table 8.

<i>GlycoMod</i> settings	
Mass value	monoisotopic
Mass tolerance	5 ppm
Ion mode	$[M+H]^+$
Form of N-linked oligosaccharide	N-X-S/T/C (X not P) motif
Enzyme	Trypsin
Max. missed cleavages	2
Cysteine treated with	Iodoacetamide
Acrylamide adducts	activated
Methionine oxidized	activated
Min. count of Hexose	3
Min. count of HexNAc	2

Table 8: *GlycoMod* software settings for glycopeptide identification:

The table contains the applied settings used for glycopeptide identification. The *GlycoMod* software is freely available under <http://web.expasy.org/glycomod/>.

3.2 Site specific N-Glycan structures in Thrombospondin-1

To evaluate the data, an Excel sheet was written for calculating possible Y-fragment masses based on the peptide masses of TSP-1. These masses were used to find accordances with the MS² spectrum of the glycopeptide. By this strategy, at least seven different glycan structures (Figure 16 – Figure 22) have been documented as being present on the same glycosylation site in TSP-1. This glycosylation site belongs to the C-terminal domain of the protein. Five of the overall seven identified glycopeptides were found being present in both, activated as well as non-activated samples. The other two glycans, both of them are of hybrid type, were found only in the IL-1 β treated sample.

The spectrum in Figure 16 shows the fragments of a hybrid type oligosaccharide containing one *N*-glycolylneuraminic acid which is linked to a peptide located in the C-term region of TSP-1. The spectrum contains B-fragment ions typical to CID fragmentation, *i.e.*, HexNAc-Hex, Hex-HexNAc-Hex, and HexNAc-Hex-Neu5Gc. These *m/z* values are indicative for *N*-linked glycopeptides. The fragment peak at *m/z* 673.160 confirms the presence of the *N*-glycolylneuraminic acid. This monosaccharide usually do not appear in human tissues due to an exon deletion in the *CMAH* gene encoding the cytosine-5'-monophospho (CMP)-Neu5Ac hydroxylase. However, Neu5Gc has often been observed in many different cancer tumours.^[91] In recent studies using a polyclonal chicken anti-Neu5Gc antibody, such glycans occasionally have also been found in healthy human tissues. It has been postulated that this monosaccharide originates from exogenic food products.^[92]

Evaluation of the Y-fragment ions indicates that α 1-3 mannose of the core is linked with two antennas whereby one of them carry the Neu5Gc. The other core mannose (α 1-6) is coupled to another mannose residue only. Most probably, the glycoprotein was transferred to the trans-Golgi before the processing by α -mannosidase II has been completed.

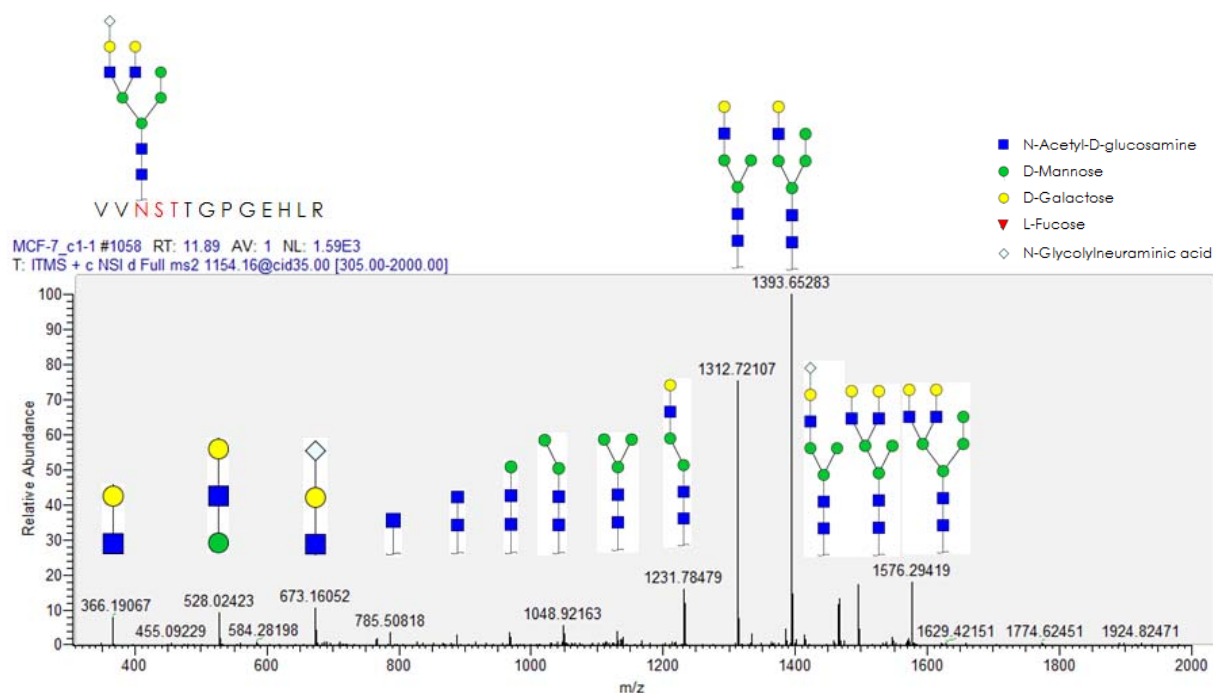


Figure 16: MS² spectrum of the indicated glycopeptide located in the C-terminal domain of TSP-1 (precursor ion: m/z 1154.16 [3⁺-charged]). The glycoprotein consists of a hybrid type oligosaccharide with one terminal *N*-Glycolylneuraminic acid. With hybrid type glycans, the α 1-6 mannose of the core carries another mannose molecule(s), whereas the α 1-3 core mannose carries both antennas.

The glycopeptides presented in Figures 17-19 contain at least one fucose unit. This monosaccharide is attached to *N*-acetylglucosamine structure only by fucosyltransferases. The first core GlcNAc (linked to asparagine) is preferred for fucosylation, however, also all GlcNAcs in the antennas can be affected by this modification.

In Figure 17 the MS² spectrum of the precursor m/z 1046.46 is shown. The fragment at m/z 512.251 represents a B-ion Hex-HexNAc-Fuc structure indicative for fucose linked to one of the antennas. Another evidence for antenna fucosylation is that a fucose containing glycan was not observed beneath Y-fragment mass of m/z 1231.969. This fragment consists of one antenna linked to mannose of the core. The next peak illustrates the first structure of the Y-series carrying fucose. Analyzing of all fragments of the Y-fragment series has allowed fitting together the bi-antennary complex structure. Though there is no fragment present which would be indicative for the presence of core-fucosylation, its absence cannot be completely excluded because of this finding.

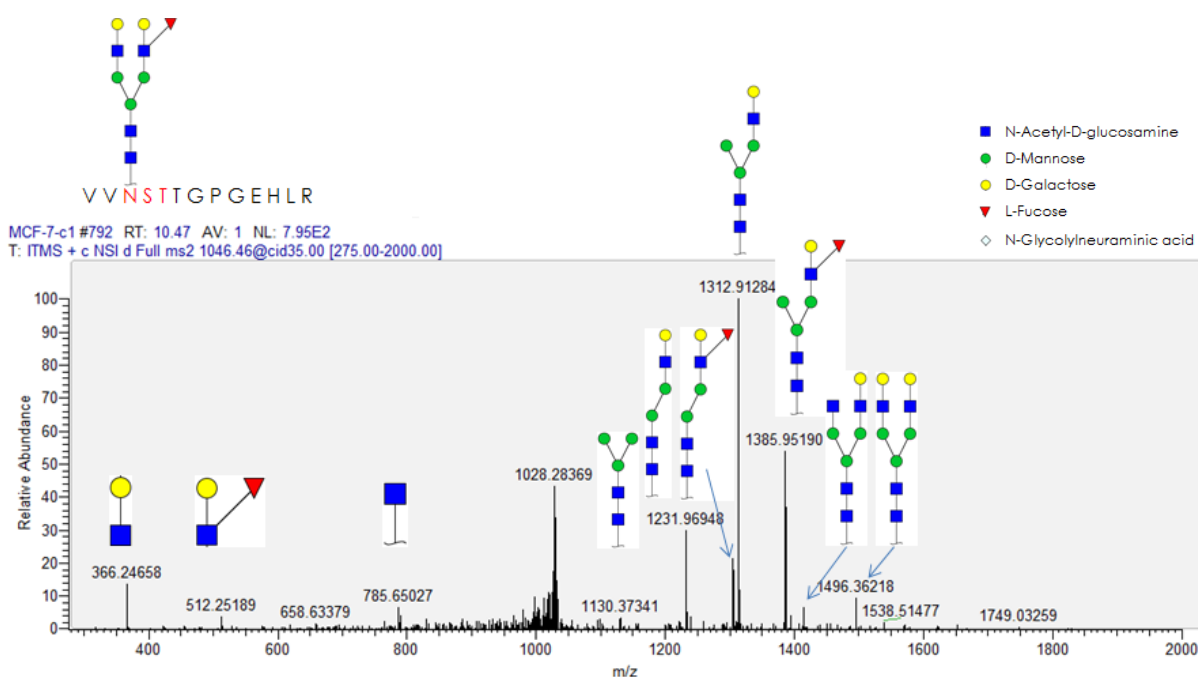


Figure 17: MS² spectrum of the indicated glycopeptide located in the C-terminal domain of TSP-1 (precursor ion: m/z 1046.46 [3⁺-charged]). The N-glycan is of complex type structure and carries one fucose molecule. The B-fragment ion at m/z 512.251 indicates that the fucose molecule is attached to the antenna. The first Y-fragment with an attached fucose molecule (m/z 1304.845) contains one antenna linked to the core mannose. Fragments of the Y-series are almost complete and prove the bi-antennary complex type structure.

The glycopeptides represented in Figure 18 and 19 exhibit the same bi-antennary glycan as in Figure 17 but in these cases here with two and three fucose units, respectively. Therefore, most of the fragments in the spectra are identical to those of Figure 17.

In Figure 18 the B-fragment ion at m/z 512.110 again is indicating for the presence of at least one fucose molecule linked to an antenna. A further fragment of the Y-series (m/z 1458.966) displays the peptide mass together with two fucoses, three GlcNAc's, and four hexoses suggesting that one fucose is attached to the core structure.

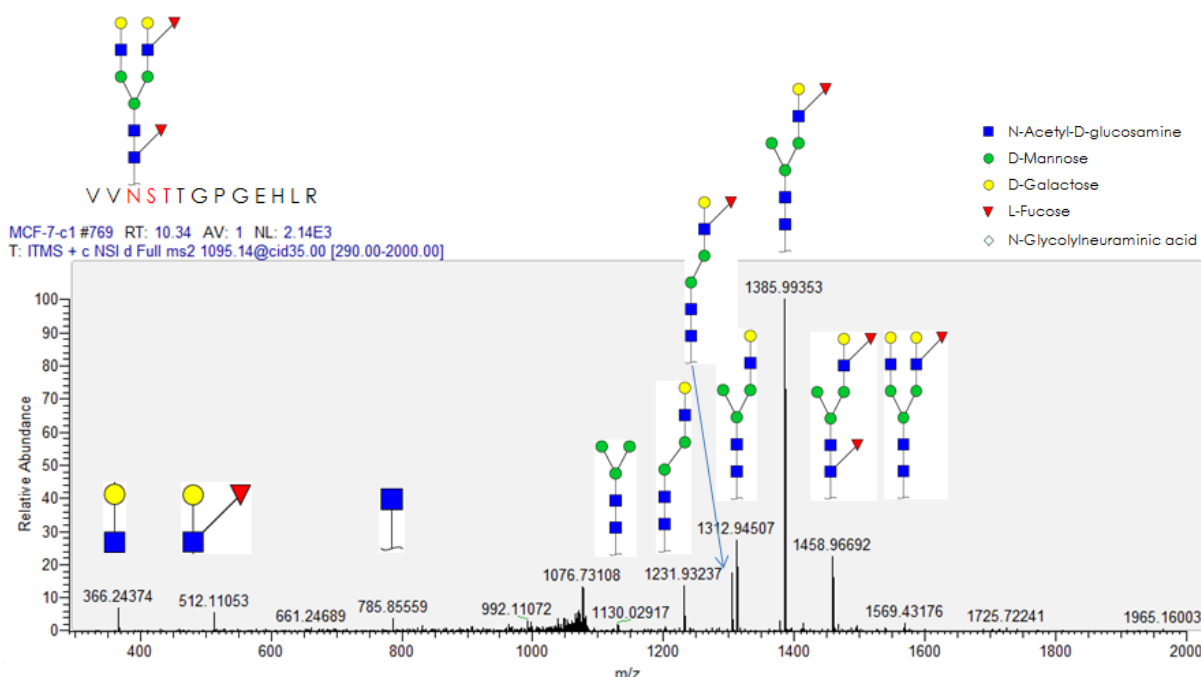


Figure 18: MS² spectrum of the indicated glycopeptide located in the C-terminal domain of TSP-1 (precursor ion: m/z 1095.14 [3⁺-charged]). The oligosaccharide is of complex type structure and di-fucosylated. The B-fragment ion at m/z 512.110 and the Y-fragment ion at m/z 1458.966 suggest that fucose molecules are attached to antenna and core, respectively. The Y-fragments are very similar to those of Figure 17 evidencing the bi-antennary glycan structure.

In Figure 19 a glycan structure with three fucoses is shown. This is an interesting spectrum because there is no clear evidence about where the third fucose is linked to. The B-ion at m/z 512.264 and the Y-ion at m/z 1459.016 confirm the attachment to an antenna and to the core. Actually, fucoses should be distributed to both antennas and to the core but the fragment next to m/z 1459.016 suggests another structure. The distance between Y-fragment ions of m/z 1459.016 and m/z 1531.936 describes an addition of one fucose indicating that probably two fucoses are attached to the GlcNAc in the core. However, such *N*-linked glycans have only been observed in invertebrates where fucoses can be linked in α 1-3 and α 1-6 position to *N*-acetylglucosamine. In such cases additional MS³ experiments might be useful to deliver more detailed information about the right position.

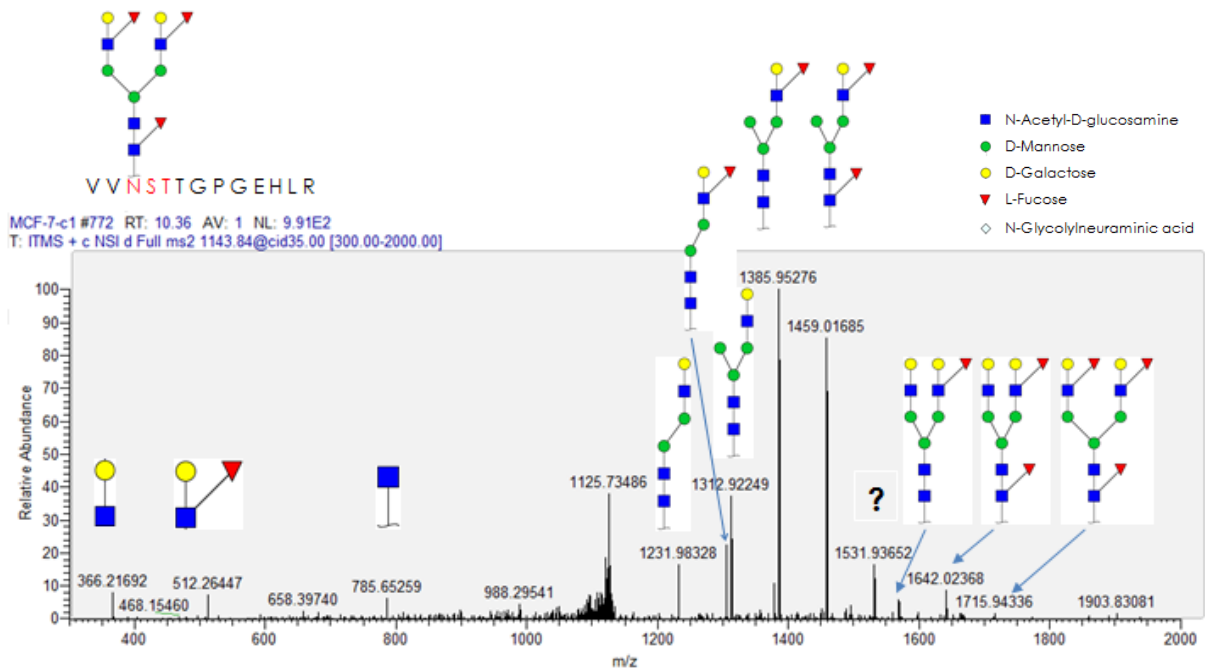


Figure 19: MS² spectrum of the indicated glycopeptide located in the C-terminal domain of TSP-1 (precursor ion: m/z 1143.84 [3⁺-charged]). The glycan structure is of complex type and tri-fucosylated. This spectrum contains no clear evidence about where the third fucose molecule is attached. The Y-fragment at m/z 1531.936 corresponds to a fragment with the structure of ion m/z 1459.016, but carrying a third fucose unit. It is not yet clear, whether two fucoses are linked to the first core GlcNAc as such a di-fucosylated core has only been observed in invertebrates so far.

The last structure of those which have been observed in both, non-treated as well as treated samples, is analyzed in Figure 20. This glycan has a hybrid type structure, with two antennas linked to the α 1-3 pentasaccharide core mannose. The other core mannose at position α 1-6 carries two mannose residues. B-ions (m/z 366.177 and m/z 528.250) of the spectrum show typical sugar fragments of antenna structures. Fragment masses derived from Y-series show a strict mass shift of a hexose molecule elucidating the high-mannose part of the glycan. Interestingly, the spectrum contains no fragment describing a structure with both antennas being of complex type.

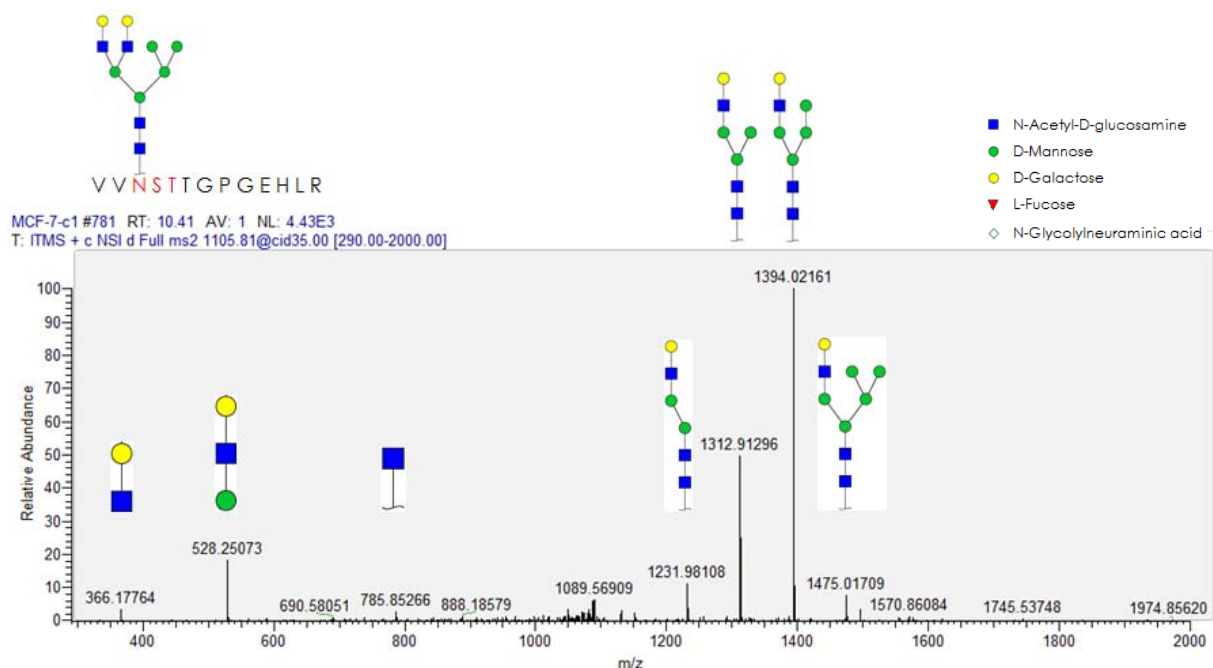


Figure 20: MS² spectrum of the indicated glycopeptide located in the C-terminal domain of TSP-1 (precursor ion: m/z 1105.81 [3⁺-charged]). The oligosaccharide is of hybrid type and contains a bi-antennary complex structure attached to the α 1-3 mannose of the core. Both mannose residues are attached to the core mannose in α 1-6 position.

The glycans described in Figure 21 and 22 have been observed in inflammatory activated (IL-1 β) samples only. Both structures are of hybrid type, with two mannose molecules attached to the Man α 1-6 arm of the core.

The glycopeptide in Figure 21 carries only one GlcNAc residue at the α 1-3 core mannose. This *N*-acetylglucosamine is added in *medial*-Golgi to initiate further mannose removal in glycan processing by α -mannosidase II. Mass differences in Y-series show all possible fragments obtained by fragmentation of linkages between mannose residues. The identified glycan usually occurs in *medial*-Golgi amongst a long procedure of glycan processing. In *trans*-Golgi processing is proceeded resulting in mature complex or hybrid type structures. The simple form of the structure found here indicates that under the given conditions (IL-1 β activation of the cells) further processing has not been applied to the glycoprotein.

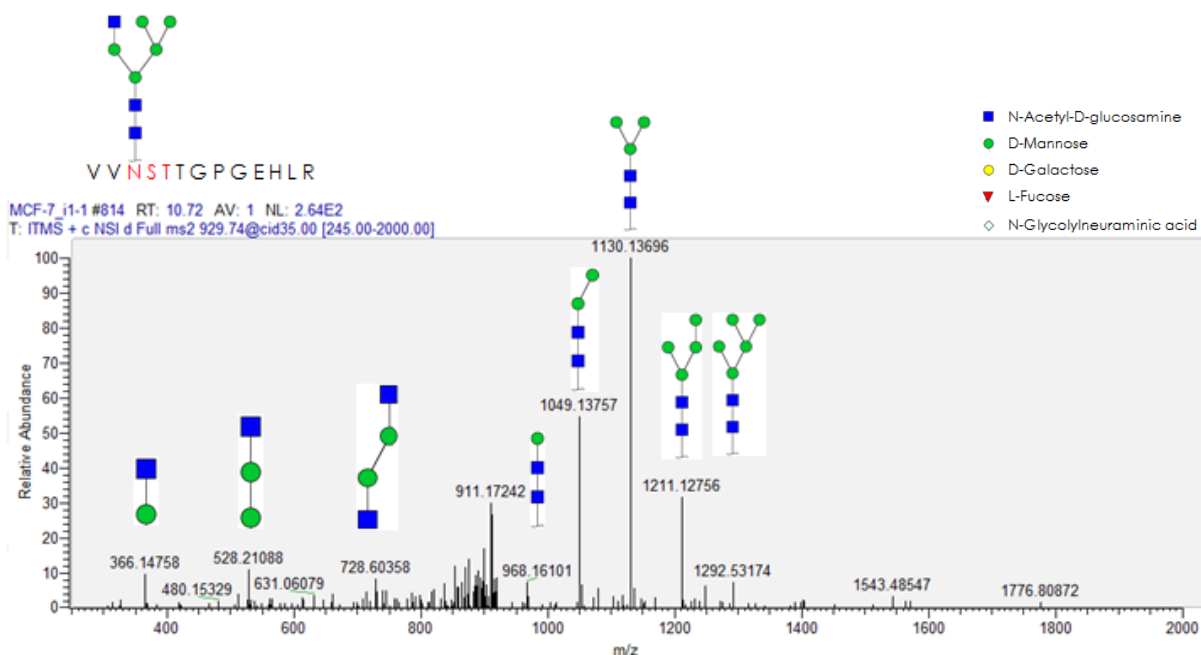


Figure 21: MS² spectrum of the indicated glycopeptide located in the C-terminal domain of TSP-1 (precursor ion: m/z 929.74 [3⁺-charged]). The *N*-glycan is of hybrid type and represents the oligosaccharide structure which occurs after glycan processing in the *medial*-Golgi. The attachment of the *N*-acetylglucosamine to the α 1-3 core mannose is mediated in *medial*-Golgi to enable further removal of mannose residues by α -mannosidase II. This *N*-glycan indicates that no further processing toward a mature hybrid or complex type structure has taken place in *trans*-Golgi.

The glycopeptide in Figure 22 is quite similar to those of Figure 20. The only difference in glycan is that the second antenna does not carry a galactosyl residue. B-ions and Y-ions are nearly identical with evaluated fragments in Figure 20. Both B-fragment ions (m/z 366.162 and m/z 528.065) display masses obtained from breaking of the antennas. Fragments of the Y-series even exhibit the same relation as in spectrum of Figure 20. Additionally, a Y-fragment (m/z 1130.150) has been observed describing the pentasaccharide core.

The spectra in Figure 20 as well as in Figure 22 do not contain Y-fragment indicating the presence of two antennas. However, such fragments have been observed in the spectrum given in Figure 16. Consequently, two explanations are possible for this phenomenon. First, the elimination of the sialic acids like *N*-glycolylneuraminic acid absorbs a lot of CID energy and limits thus further fragmentations of antenna. A second postulation describes an effect which arises from the two mannose residues

coupled to Man α 1-6 arm of the core. So maybe these mannoses have some steric effect on complex antennas that one of them will more easily be dissociated by collisions with gas molecules.

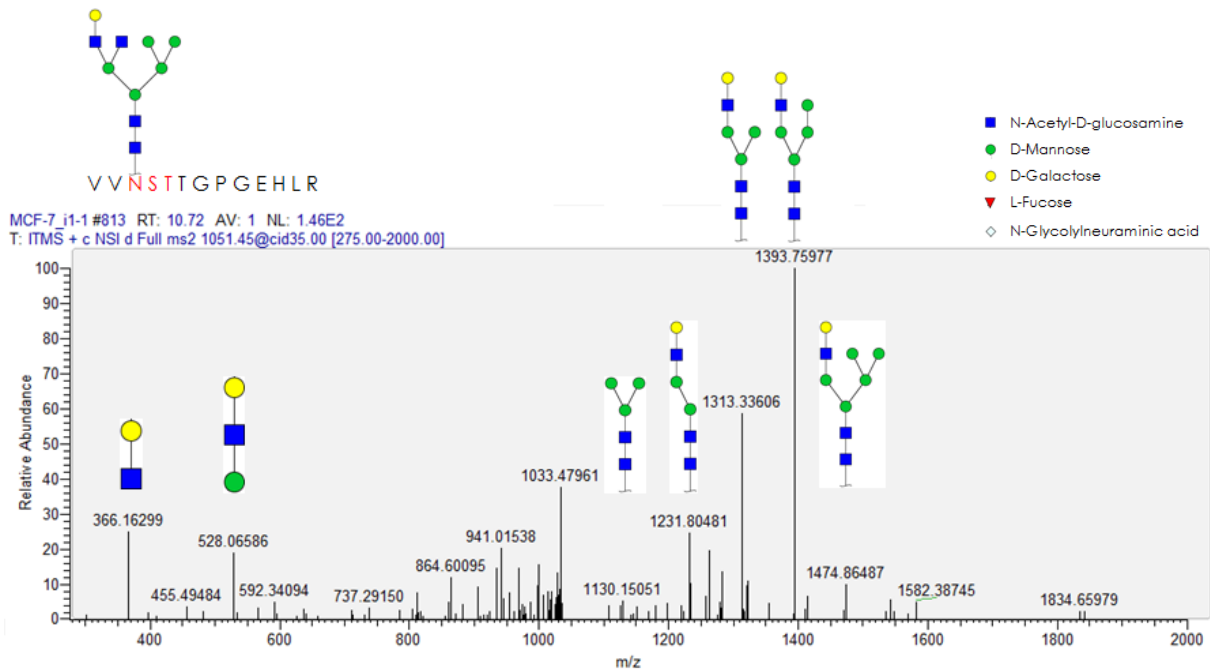


Figure 22: MS² spectrum of the indicated glycopeptide located in the C-terminal domain of TSP-1 (precursor ion: m/z 1051.45 [3⁺-charged]). The oligosaccharide is of hybrid type containing two complex antennas (one of them without a galactosyl residue) and two mannose residues. With hybrid type glycans, the antennas are linked to the α 1-3 core mannose, the additional mannose residues to the α 1-6 core mannose.

3.3 ETD Spectrum of Glycopeptides

Figure 23 displays an ETD spectrum typical for one of the identified glycopeptides of TSP-1. The spectrum was obtained from the peptide carrying a hybrid type *N*-glycan with precursor mass m/z 1105.47 (3+ charged). The figure illustrates that ETD fragmentation of glycopeptides results mainly in c- and z-fragments of low intensities which mostly appear at m/z values below the precursor ion. This phenomenon has also been observed in former investigations by Mayampurath *et al.*^[83]

66

Measurements with bovine alpha-1-acid glycoprotein (AGP) as a standard protein have shown that the higher loaded the precursor ion is, the better will be the relative abundance of fragment ions (Supplemental Material). To get a clearer overview about all evaluated fragments, the original spectrum (displayed at the bottom) is zoomed-in (displayed at the top) to highlight the region between m/z 200 and the precursor ion (m/z 1105.438).

The zoomed-in spectrum shows all identified fragments obtained from the peptide backbone. All these fragments are from the z-series between numbers three and ten with exception of z_6 -ion. Responsible for the absence of this fragment is the fact that breaking of the C-N bond in proline does not result in separated fragments due to the cyclic structure of proline. No fragments of the c-series have been observed because the *N*-glycan is linked to the third amino acid in the peptide sequence. Therefore, possible c-fragments would exceed the mass range (m/z 0 to 2,000) due to the high mass of the glycan structure.

The occurrence of the Hex-HexNAc fragment ion (characterised peak at m/z 366.194) is highly interesting because such a glycan fragment does not appear by ETD fragmentation. Usually, ETD leads to c- and z-fragment ions described in Figure 8. However, other measurements involving bovine AGP have also shown such fragments originated from complex glycan antennas (Supplemental Material). An explanation for this phenomenon might be that these linkages in glycans are really weak and therefore they get easily activated through transport processes within the instrument. In further consequence, the bonds will be broken automatically where

the electron transfer process takes place. ETD fragmentation has shown to be a good method for getting information about the peptide backbone.

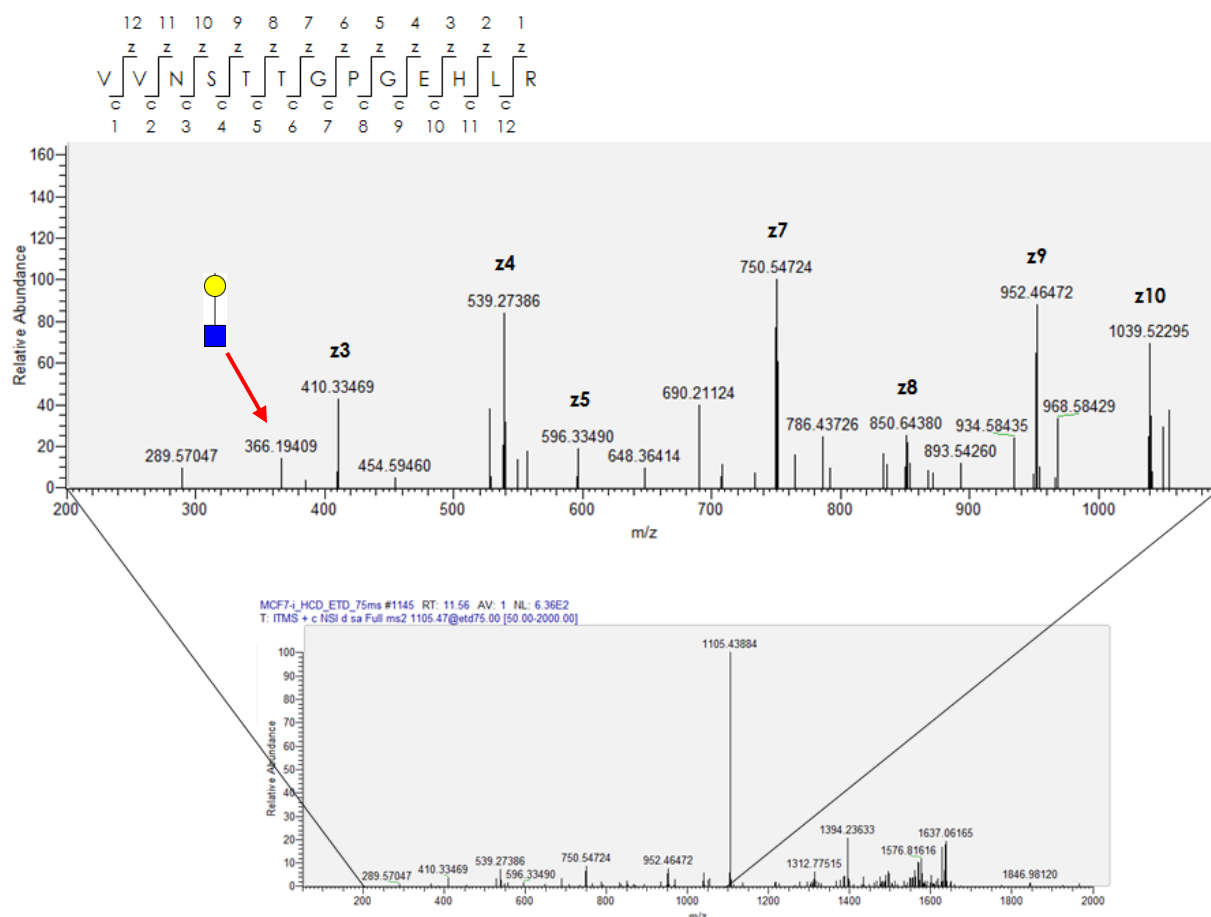


Figure 23: ETD fragment spectrum of a glycopeptide located in the C-terminal domain of TSP-1 (precursor ion: m/z 1105.81 [3⁺-charged]). The spectrum at the top shows the zoomed-in region between m/z 200 and the m/z value of the precursor. Identified z-fragment ions are indicated. No c-fragment ions have been observed because the *N*-glycan is attached to the third amino acid and, therefore, masses of these ions lie outside the m/z range. The peak indicated by the red arrow originates from the Hex-HexNAc ion. Its occurrence in an ETD spectrum is interesting as ETD fragmentation usually does not result in B-fragment ions.

3.4 Relative Quantification of different Glycopeptides

Figure 24 shows relative quantification of all identified *N*-glycan species in TSP-1 present in the SN of MCF-7 cells without or after inflammatory stimulation. Average values and standard deviations characterizing the technical reproducibility were calculated on the base of two measurements. For this quantification, the glycopeptide mass which exhibited the highest number of ions in the MS¹ spectrum was set to 100%. Subsequently, amounts of all other glycopeptides were normalized to this most abundant glycopeptide. In the diagram, relative amounts (y-axis) are given for the glycopeptides with various oligosaccharide structures characterized by their precursor masses (x-axis). On the bottom of the figure precursor masses are described by picturing the glycan structures which are linked to the same *N*-site in TSP-1.

In both non-activated as well as IL-1 β activated sample, the glycopeptide with precursor mass m/z 1105.47 has reached the highest abundance. Interestingly, this glycan is of hybrid type. The glycopeptide containing the *N*-glycolylneuraminic acid is the second most abundant structure with about 50%. Percentage values for complex types possessing fucose residues lie in the range of 10-35%. The last two glycopeptides showed in the diagram have only been observed in IL-1 β treated sample. These compounds have reached an abundance of about 10% compared to the most common one.

The differences seen between control and IL-1 β activated samples are of sincere interest, though the number of biological repeats is at the moment too low to come to a firm conclusion. However, there are some preliminary conclusions which have to be corroborated by a larger number of samples. Glycopeptides with the Neu5Gc residue exhibited no variation in abundance when comparing activated vs. control samples, but with glycopeptides carrying complex type oligosaccharide structures certain variations in abundance were observed. Particularly, the relative amount of the di- and trifucosylated oligosaccharides attached to the peptides is much lower in the IL-1 β activated sample compared to control. And this is in contrast to the glycopeptide with monofucosylated glycan, which is twice as abundant than found in the control.

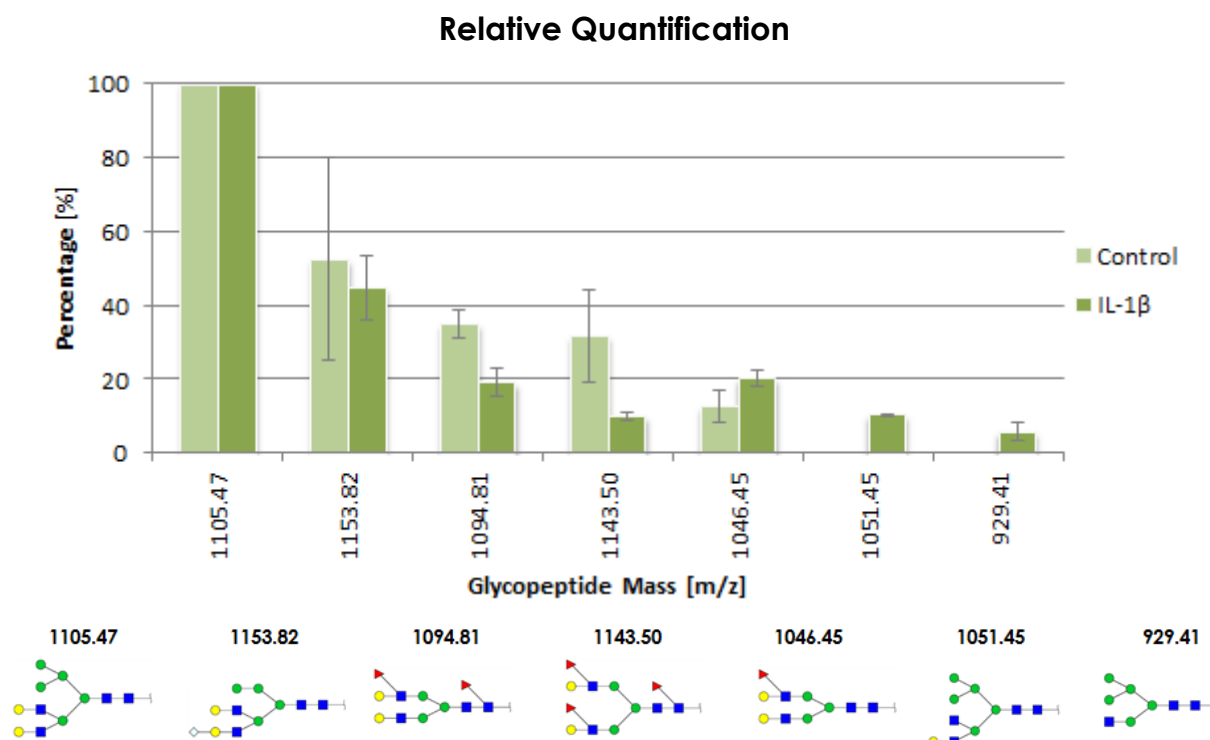


Figure 24: Relative abundances of all identified glycopeptides from TSP-1.

The diagram compares relative abundances of glycopeptide species of the peptide VVNSTGPGEHLR carrying different oligosaccharide structures in control and IL-1 β activated sample. Amounts are normalized to that of the most abundant species. Glycan structures corresponding to the glycopeptide masses are pictured on the bottom

Identification of glycopeptides in control as well as IL-1 β treated sample has not shown a serious alteration (e.g. introduction of a bisecting GlcNAc) in glycan structure. However, relative quantification displays some differences in abundances of certain glycopeptides. It seems that, lesser processed *N*-glycans (e.g. monofucosylated type or incomplete hybrid types) are more abundant in inflammatory activated sample than in the control. Especially, glycan processing realised in the *trans*-Golgi apparatus seems to be affected by this IL-1 β treatment. It might be an indication that the enzymatic machinery responsible for assembly of mature *N*-glycans is disturbed by the activation by IL-1 β .

4. Conclusion

In this master thesis MS-based strategies were investigated for monitoring alterations in *N*-linked glycosylation patterns upon cell activation. SN of MCF-7 cells (non-activated and IL-1 β activated) were used to evaluate and optimize the preparation procedure and MS settings. For suitable glycoprotein analysis, two MS/MS methods have been established for elucidating the structures of the glycan species as well as those of the peptide backbone.

70

The first strategy for glycopeptide analysis was the development of a Top6 method with CID fragmentation. Thereby, the six most abundant peaks in the MS¹ spectrum were used for following fragmentation applying CID. With glycopeptides, this technique predominantly leads to B- and Y-fragment ions that deliver information about the glycan attached to the peptide. Fragments of B-series (particularly Hex-HexNAc peak) can be taken to decide whether the spectrum is a glycopeptide or not. As a result, glycopeptide mass is available for searching possible structures via databases. Analyses of B- and Y-fragments have shown some certain fragmentation patterns typically obtained by CID. For instance, the chemical bond between a core mannose and the GlcNAc of the antenna is most commonly fragmented by CID. Moreover, fragmentation of glycopeptides containing fucoses indicated that linkages between GlcNAc and fucose will hardly be dissociated. A fragment often present after CID is the peptide-GlcNAc ion. Five out of the seven declared CID spectra contain this Y-fragment which simply allows calculating the peptide mass.

Another method applying a data dependent HCD and ETD fragmentation was developed to acquire knowledge about the peptide backbone. In this setting, an ETD event will be triggered only if HCD fragmentation leads to certain B-fragment ions. This strategy guarantees an exclusive utilization of ETD for the fragmentation of glycopeptides. CID fragmentation would also be applicable instead of HCD to trigger a following ETD event. However, fragments of HCD are analyzed in the Orbitrap leading to a higher accuracy and therefore the ETD triggering process would be more specific.

The electron transfer process yields c- and z-fragments of the peptide backbone and, thus, ETD has become the method of choice for analyzing peptide sequences carrying PTMs. However, in our case some B-fragment ions have been identified which originate from *N*-glycan. All of them were typical fragments of glycan antennas and might have some other reasons beside the ETD fragmentation process. One explanation could be that antennas get unintentionally collision activated in some way and will dissociate during ETD processing.

ETD spectra show small relative abundances (<10%) of the fragments compared to the precursor ion. Nevertheless, if the region beneath the precursor mass is zoomed-in, nearly the entire fragment series is identifiable. In case of proline, the ETD fragmentation of the c/z bond does not result in separated fragments.

To sum up, separation techniques such as HPLC coupled to ESI-MS is a powerful strategy for fast analysis of many glycoproteins in complex samples. Therefore MS delivers several fragmentation methods to acquire different information about the structures. For future investigations in glycoproteomics issues, efforts are necessary to facilitate low abundant protein analysis. Normally, in body fluids a few proteins cover nearly the whole amount of proteins making glycoproteomic studies difficult. However, structure elucidation of low expressed proteins would have been important in disease research. At the moment, there is no adequate evaluation software available for automatic glycopeptide identification. Such an algorithm would be helpful for saving time which will be needed in developing analytical applications. Hence, there is still place for further efforts to push glycoproteomics studies to the next level.

5. Supplemental Material

Data dependent Top6 method with HCD and ETD fragmentation was evaluated by measuring bovine AGP as standard protein. The results are shown in Figure 25 where base peak chromatograms (BPC) of full scan and from fragmentation with HCD and ETD are presented. The chromatogram on the top (black colour) constitutes the full scan BPC which was acquired in Orbitrap analyzer with a resolution of 15,000. Precursor ions of the six most intense peaks had been fragmented with HCD technology shown by the red coloured chromatogram in the middle of the figure. To reach a high accuracy, fragment ions obtained from HCD were also acquired in the Orbitrap. This is necessary because the decision whether an ETD event will be triggered or not depends on certain peaks in HCD spectrum. So, if a HCD fragmentation spectrum contains the HexNAc (m/z 204.087) and the Hex-HexNAc (m/z 366.138) peaks within the ten most intense peaks, an ETD fragmentation is triggered on the same precursor ion. The BPC (green colour) on the bottom illustrates a fragmentation spectrum of ETD events.

The evaluation of data shows that the method works successfully. ETD fragmentation had only been applied to those precursor ions which exhibit the typical two oxonium ion peaks by HCD.

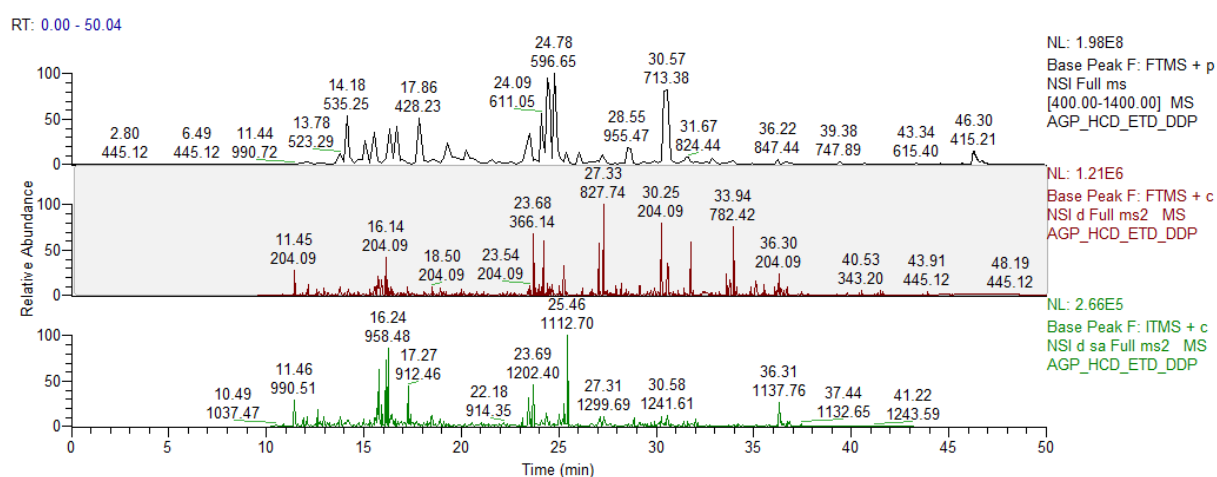


Figure 25: Base peak chromatogram (BPC) obtained by applying the data dependent Top6 method on bovine AGP. The chromatogram of the full scan and the MS² BPC of the HCD fragmentation are shown in black and red colour, respectively. MS² BPC of the ETD fragmentation is represented by the green

colour. This fragmentation was only triggered if the HCD fragmentation delivered HexNAc (m/z 204.087) and Hex-HexNAc (m/z 366.138) fragments within the ten most intense peaks.

Figure 26 represents a MS^2 spectrum usually obtained by HCD fragmentation applied to glycopeptides. Characterised peaks (red arrows) highlight the HexNAc (m/z 204.086) and the Hex-HexNAc (m/z 366.139) fragment peaks which are necessary to trigger an ETD fragmentation on the same precursor ion. However, intensity of these peaks has to be among the ten most abundant peaks. If HCD fragmentation yields only one of these B-fragments or peaks are too small no extra ETD fragmentation will be triggered.

73

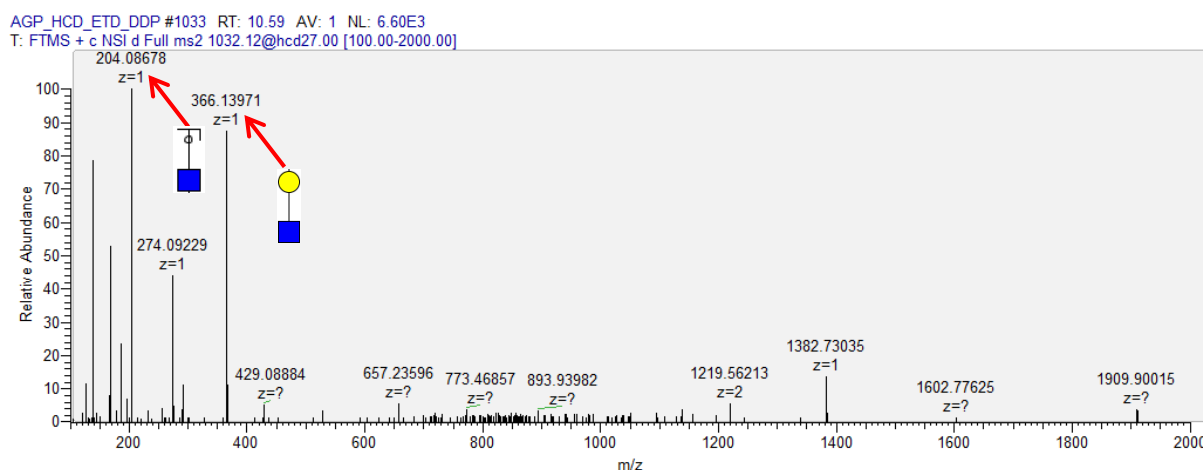


Figure 26: HCD fragment spectrum of an *N*-linked glycopeptide from the standard protein bovine AGP. The peaks indicated by red arrows originate from HexNAc (m/z 204.086) and Hex-HexNAc (m/z 366.139) ions. Both peaks have to appear within the ten most abundant peaks to trigger an ETD event on the same precursor ion.

The electron transfer process ETD fragmentation primarily results in c- and z- fragment ions of peptide backbone. Therefore, this technique becomes popular for analyzing post translational modifications such as glycosylation or phosphorylation. Contrary to CID fragmentation, ETD should result in fragment ions which still have the intact sugar chain attached. However, experiments applying ETD fragmentation have shown that there appear fragment ions originating from glycan structures particularly from antennae. Figure 27 presents a spectrum of ETD fragmentation in the Orbitrap Velos-instrument showing typical oxonium fragment ions from glycan structures. The char-

acterised peaks originated from the Hex-HexNAc (m/z 366.316), the Hex-HexNAc-Neu5Ac (m/z 657.234) and the Hex-HexNAc-Neu5Gc (m/z 673.293) fragment ions. The occurrence of such sugar fragments in this intensity indicates that the chemical bond between the core mannose and the HexNAc of the antenna must be weak.

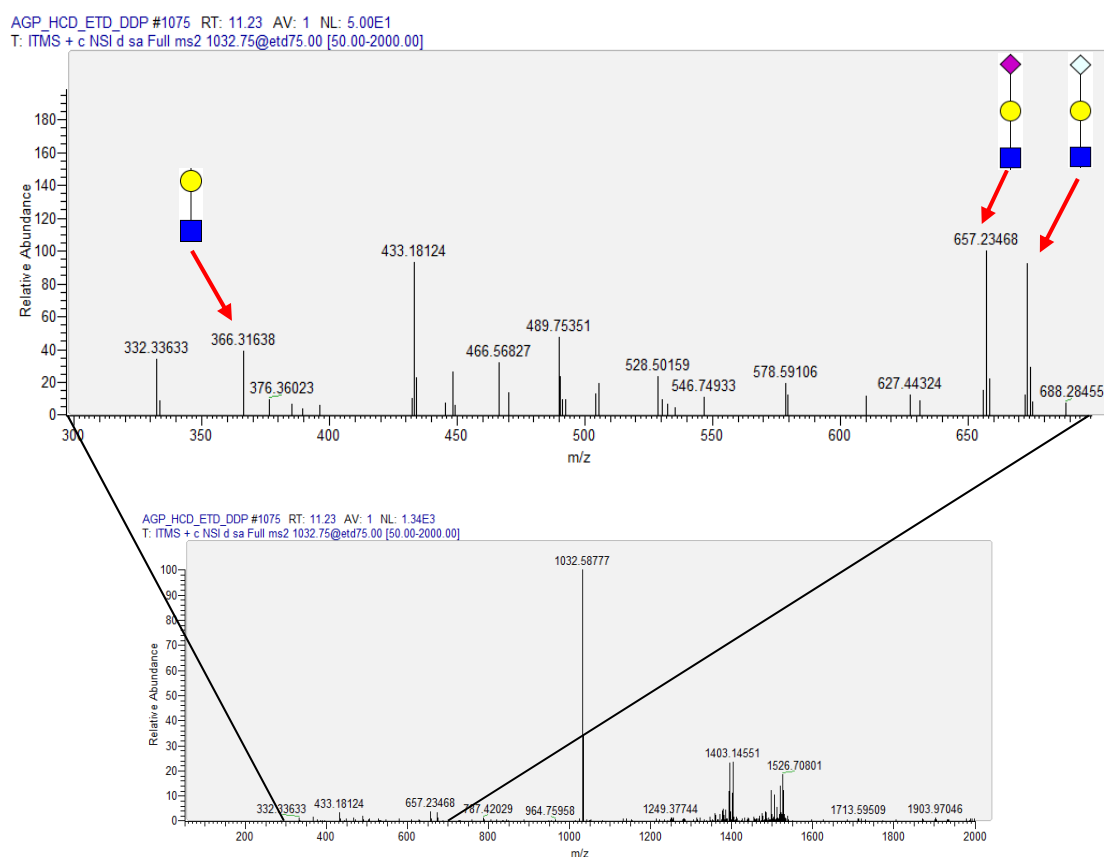
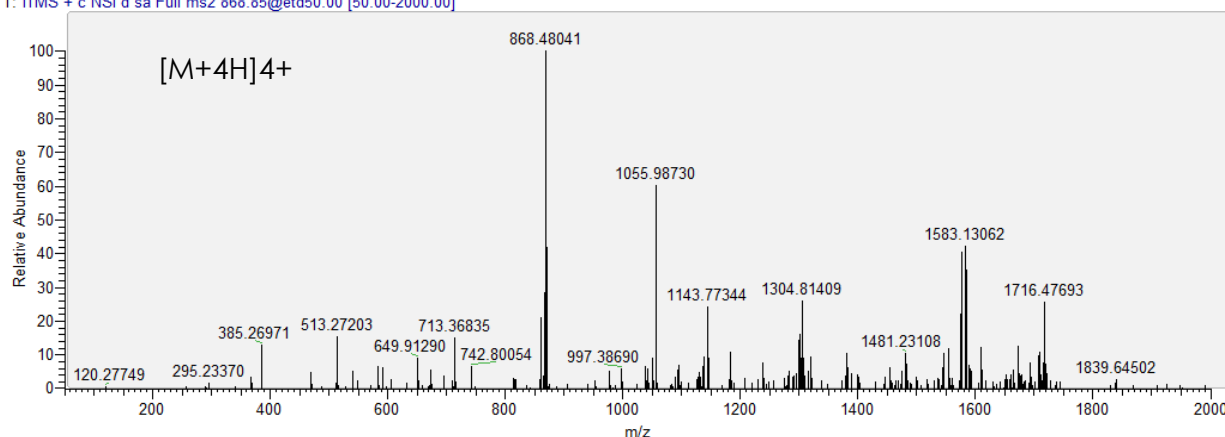


Figure 27: ETD fragment spectrum of a bovine AGP glycopeptide with precursor m/z 1032.75. The peaks indicated by red arrows originate from B-fragment ions which are untypical for ETD fragmentation. PTMs such as glycosylation usually should be unaffected by ETD fragmentation. However, such fragment ions originate from di- and trisaccharide molecules of the *N*-glycan have been observed in ETD spectra many times.

Spectra shown in Figure 28 illustrate the difference of ETD fragmentation applied to a glycopeptide whereby the precursor ion was triply (spectrum on the bottom) and four-fold (spectrum on the top) charged, respectively. The settings of the method and the conditions of acquisition were the same in both cases. The differences between both spectra are noticeable at a glance. Fragmentation of the four-fold loaded precursor ion yields a higher number of fragments than the fragmentation of

the triply charged one. Both, the peak intensity and the number of obtained peaks are higher in the spectrum on the top. Reasonable for this quality enhancement in higher charged precursor ions is the mechanism behind the electron transfer process. Higher charged molecules more easily attract the electron from fluoranthene reagent ion. Therefore, the transfer process is faster and leads to a higher quality of the fragment spectra.

AGP_CID_ETD_50ms #724 RT: 10.82 AV: 1 NL: 3.92E2
T: ITMS + c NSI d sa Full ms2 868.85@etd50.00 [50.00-2000.00]



AGP_CID_ETD_50ms #711 RT: 10.72 AV: 1 NL: 5.22E2
T: ITMS + c NSI d sa Full ms2 1158.46@etd50.00 [50.00-2000.00]

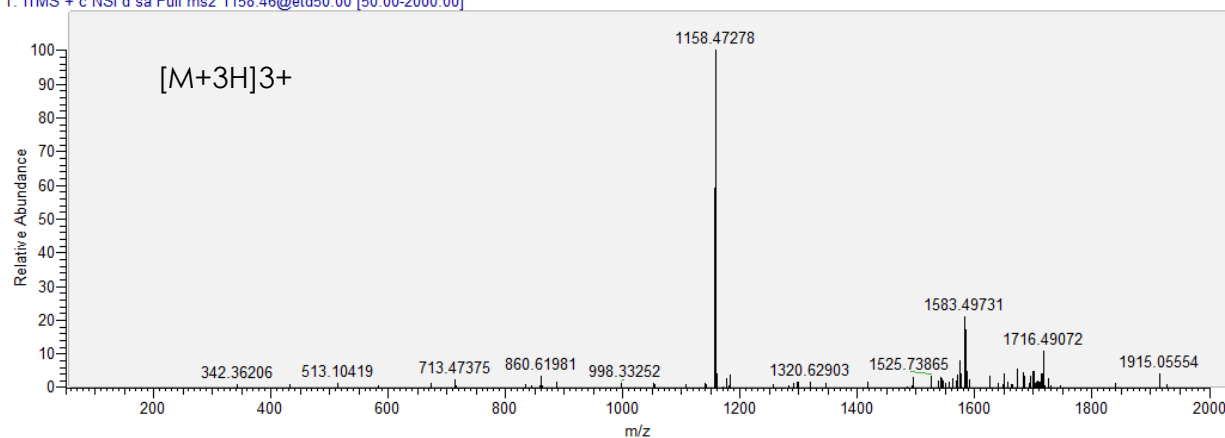


Figure 28: Comparison of ETD fragmentation applied on the same precursor (precursor mass: 3473.387) but with different charge states. The fragment spectrum on the top (4⁺-charged precursor ion) exhibits many more peaks and higher intensity of fragments compared to the fragment spectrum on the bottom (3⁺-charged precursor). This indicates that the higher loaded the precursor ion is the higher is the quality of the ETD fragment spectrum.

VI. REFERENCES

- 1) Spiro RG; Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds; *Glycobiology* **2002**; 12: 43R-56R.
- 2) Weerapana E, Imperiali B; Asparagine-linked protein glycosylation: from eukaryotic to prokaryotic systems; *Glycobiology* **2006**; 16: 91R-101R.
- 3) Van den Steen P, Rudd PM, Dwek RA, Opdenakker G; Concepts and principles of O-linked glycosylation; *Crit. Rev. Biochem. Mol. Biol.* **1998**; 33: 151-208.
- 4) de Beer T, Vliegthart JFG, Löffler A, Hofsteenge J; The Hexapyranosyl Residue That Is C-Glycosidically Linked to the Side Chain of Tryptophan-7 in Human RNase U_s Is α -Mannopyranose; *Biochemistry* **1995**; 34: 11785-11789.
- 5) Haynes PA; Phosphoglycosylation: a new structural class of glycosylation?; *Glycobiology* **1998**; 8: 1-5.
- 6) Thomas JR, Dwek RA, Rademacher TW; Structure, Biosynthesis, and Function of Glycosylphosphatidylinositols; *Biochemistry* **1990**; 29: 5413-5422.
- 7) Marshall RD; The nature and metabolism of carbohydrate peptide linkages of glycoproteins; *Biochem. Soc. Symp.* **1974**; 40: 17-26.
- 8) Bause E; Structural requirements of N-glycosylation of proteins. Studies with proline peptides as conformational probes; *Biochem. J.* **1983**; 209: 331-336.
- 9) Valliere-Douglass JF, Eakin CM, Wallace A, Ketchum RR, Wang W, Treuheit MJ, Balland A; Glutamine-linked and Non-consensus Asparagine-linked Oligosaccharides present in Human Recombinant Antibodies Define Novel Protein Glycosylation Motifs; *J. Biol. Chem.* **2010**; 285: 16012-16022.
- 10) Satomi Y, Shimonnishi Y, Takao T; N-glycosylation at Asn⁴⁹¹ in the Asn-Xaa-Cys motif of human transferrin; *FEBS Lett.* **2004**; 576: 51-56.
- 11) Stanley P, Schachter H, Taniguchi N; N-Glycans In: *Essentials of Glycobiology, Second Edition*; Cold Spring Harbor Laboratory Press, New York, **2009**.
- 12) Aebersold M; N-linked protein glycosylation in the ER; *Biochim. Biophys. Acta* **2013**; 1833: 2430-2437.
- 13) Schenk B, Fernandez F, Waechter CJ; The ins(die) and outs(die) of dolichyl phosphate biosynthesis and recycling in the endoplasmic reticulum; *Glycobiology* **2001**; 11: 61R-70R.

- 14) Bickel T, Lehle L, Schwarz M, Aebi M, Jakob CA; Biosynthesis of Lipid-linked Oligosaccharides in *Saccharomyces cerevisiae*: Alg13p and Alg14p form a complex required for the formation of GlcNAc(2)-PP-dolichol; *J. Biol. Chem.* **2005**; 280: 34500-34506.
- 15) Gao XD, Nishikawa A, Dean N; Physical interactions between the Alg1, Alg2, and Alg11 mannosyltransferases of the endoplasmic reticulum; *Glycobiology* **2004**; 14: 559-570.
- 16) Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, Hart GW, Etzler ME; *Essentials of Glycobiology, Second Edition*; Cold Spring Harbor Laboratory Press, New York, **2009**.
- 17) Aebi M, Gassenhuber J, Domdey H, te Heesen S; Cloning and characterization of the ALG3 gene of *Saccharomyces cerevisiae*; *Glycobiology* **1996**; 6: 439-444.
- 18) Reiss G, te Heesen S, Zimmerman J, Robbins PW, Aebi M; Isolation of the ALG6 locus of *Saccharomyces cerevisiae* required for glucosylation in the N-linked glycosylation pathway; *Glycobiology* **1996**; 6: 493-498.
- 19) Freeze HH, Aebi M; Altered glycan structures: the molecular basis of congenital disorders of glycosylation; *Curr. Opin. Struct. Biol.* **2005**; 15: 490-498.
- 20) Wacker M, Linton D, Hitchen PG, Nita-Lazar M, Haslam SM, North SJ, Panico M, Morris HR, Dell A, Wren BW, Aebi M; N-Linked Glycosylation in *Campylobacter jejuni* and Its Functional Transfer into *E. coli*; *Science* **2002**; 298: 1790-1793.
- 21) Kelleher DJ, Gilmore R; An evolving view of the eukaryotic oligosaccharyl-transferase; *Glycobiology* **2006**; 16: 47R-62R.
- 22) Ruddock LW, Molinari M; N-glycan processing in ER quality control; *J. Cell Sci.* **2006**; 119: 4373-4380.
- 23) Kornfeld R, Kornfeld S; Assembly of asparagine-linked oligosaccharides; *Ann. Rev. Biochem.* **1985**; 54: 631-664.
- 24) Varki A, Lowe JB; Biological Roles of Glycans In: *Essentials of Glycobiology, Second Edition*; Cold Spring Harbor Laboratory Press, New York, **2009**.
- 25) Sinclair AM, Elliott S; Glycoengineering: The Effect of Glycosylation on the Properties of Therapeutic Proteins; *J. Pharm. Sci.* **2005**; 94: 1626-1635.
- 26) Gu J, Taniguchi N; Potential of N-glycan in cell adhesion and migration as either a positive or negative regulator; *Cell Adh. Migr.* **2008**; 2: 243-245.

- 27) Frost DC, Li L; Recent Advances in Mass Spectrometry-Based Glycoproteomics; *Adv. Protein Chem. Struct. Biol.* **2014**; 95: 71-123.
- 28) Hua S, Saunders M, Dimapasoc LM, Jeong SH, Kim BJ, Kim S, So M, Lee KS, Kim JH, Lam KS, Lebrilla CB, An HJ; Differentiation of Cancer Cell Origin and Molecular Subtype by Plasma Membrane N-Glycan Profiling; *J. Proteome Res.* **2014**; 13: 961-968.
- 29) Drake PM, Schilling B, Niles RK, Prakobphol A, Li B, Jung K, et al; Lectin Chromatography/Mass Spectrometry Discovery Workflow Identifies Putative Biomarkers of Aggressive Breast Cancers; *J. Proteome Res.* **2012**; 11: 2508-2520.
- 30) Fodero LR, Sáez-Valero J, Barquero MS, Marcos A, McLean CA, Small DH; Wheat germ agglutinin-binding glycoproteins are decreased in Alzheimer's disease cerebrospinal fluid; *J. Neurochem.* **2001**; 79: 1022-1026.
- 31) Schedin-Weiss S, Winblad B, Tjernberg LO; The role of protein glycosylation in Alzheimer disease; *FEBS J.* **2014**; 281: 46-62.
- 32) Anderson NL, Anderson NG; The Human Plasma Proteome; *Mol. Cell. Proteomics* **2002**; 1: 845-867.
- 33) Hage DS, Anguizola JA, Bi C, Li R, Matsuda R, Papastavros E, Pfaunmiller E, Vargas J, Zheng X; Pharmaceutical and biomedical applications of affinity chromatography: Recent trends and developments; *J. Pharm. Biomed. Anal.* **2012**; 69: 93-105.
- 34) Moser AC, Hage DS; Immunoaffinity chromatography: an introduction to applications and recent developments; *Bioanalysis* **2010**; 2: 769-790.
- 35) Cellar NA, Karnoup AS, Albers DR, Langhorst ML, Young SA; Immunodepletion of high abundance proteins coupled on-line with reversed-phase liquid chromatography: A two-dimensional LC sample enrichment and fractionation technique for mammalian proteomics; *J. Chromatogr. B* **2009**; 877: 79-85.
- 36) Kullolli M, Hancock WS, Hincapie M; An Automated Platform for Fractionation of Human Plasma Glycoproteome in Clinical Proteomics; *Anal. Chem.* **2010**; 82: 115-120.
- 37) Boschetti E, Righetti PG; Low-abundance proteome discovery: state of the art and protocols; Elsevier, Amsterdam, **2013**; 1-342.

- 38) Righetti PG; The Monkey King: A personal view of the long journey towards a proteomic Nirvana; *J. Proteomics* **2014**; 107: 39-49.
- 39) Zhang H, Li XJ, Martin DB, Aebersold R; Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry; *Nat. Biotechnol.* **2003**; 21: 660-666.
- 40) Xu Y, Wu Z, Zhang L, Lu H, Yang P, Webley PA, Zhao D; Highly Specific Enrichment of Glycopeptides Using Boronic Acid-Functionalized Mesoporous Silica; *Anal. Chem.* **2009**; 81: 503-508.
- 41) Kaji H, Saito H, Yamauchi Y, Shinkawa T, Taoka M, Hirabayashi J, Kasai K, Takahashi N, Isobe T; Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins; *Nat. Biotechnol.* **2003**; 21: 667-672.
- 42) Demelbauer UM, Plematl A, Josic D, Allmaier G, Rizzi A; On the variation of glycosylation in human plasma derived antithrombin; *J. Chromatogr. A* **2005**; 1080: 15-21.
- 43) Wang Y, Wu S, Hancock WS; Approaches to the study of N-linked glycoproteins in human plasma using lectin affinity chromatography and nano-HPLC coupled to electrospray linear ion trap - Fourier transform mass spectrometry; *Glycobiology* **2006**; 16: 514-523.
- 44) Fanayan S, Hincapie M, Hancock WS; Using lectins to harvest the plasma/serum glycoproteome; *Electrophoresis* **2012**; 33: 1746-1754.
- 45) West I, Goldring O; Lectin Affinity Chromatography; *Methods Mol. Biol.* **2004**; 244: 159-166.
- 46) Cummings RD, Kornfeld S; Fractionation of Asparagine-linked Oligosaccharides by Serial Lectin-Agarose Affinity Chromatography. A rapid, sensitive, and specific technique; *J. Biol. Chem.* **1982**; 257: 11235-11240.
- 47) Yang Z, Hancock WS; Approach to the comprehensive analysis of glycoproteins isolated from human serum using a multi-lectin affinity column; *J. Chromatogr. A* **2004**; 1053: 79-88.
- 48) Lundquist JJ, Toone EJ; The Cluster Glycoside Effect; *Chem. Rev.* **2002**; 102: 555-578.
- 49) Griffiths J; A Brief History of Mass Spectrometry; *Anal. Chem.* **2008**; 80: 5678-5683.

- 50) Karas M, Bachmann D, Hillenkamp F; Influence of the Wavelength in High-Irradiance Ultraviolet Laser Desorption Mass Spectrometry of Organic Molecules; *Anal. Chem.* **1985**; 57: 2935-2939.
- 51) Yamashita M, Fenn JB; Electrospray Ion Source. Another Variation on the Free-Jet Theme; *J. Phys. Chem.* **1984**; 88: 4451-4459.
- 52) Stutz H; Advances in the analysis of proteins and peptides by capillary electrophoresis with matrix-assisted laser desorption/ionization and electrospray-mass spectrometry detection; *Electrophoresis* **2005**; 26: 1254-1290.
- 53) Makarov A, Denisov E, Kholomeev A, Balschun W, Lange O, Strupat K, Horning S; Performance Evaluation of a Hybrid Linear Ion Trap/Orbitrap Mass Spectrometer; *Anal. Chem.* **2006**; 78: 2113-2120.
- 54) Ho CS, Lam CWK, Chan MHM, Cheung RCK, Law LK, Lit LCW, Ng KF, Suean MWM, Tai HL; Electrospray Ionisation Mass Spectrometry: Principles and Clinical Applications; *Clin. Biochem. Rev.* **2003**; 24: 3-12.
- 55) Taylor G; Disintegration of Water Drops in an Electric Field; *Proc. R. Soc. Lond. A* **1964**; 280: 383-397.
- 56) Bruins AP; Mechanistic aspects of electrospray ionization; *J. Chromatogr. A* **1998**; 794: 345-357.
- 57) Dole M, Mack LL, Hines RL, Mobley RC, Ferguson LD, Alice MB; Molecular Beams of Macroions; *J. Chem. Phys.* **1968**; 49: 2240-2249.
- 58) Iribarne JV, Thomson BA; On the evaporation of small ions from charged droplets; *J. Chem. Phys.* **1976**; 64: 2287-2294.
- 59) Wilm M, Mann M; Analytical Properties of the Nanoelectrospray Ion Source; *Anal. Chem.* **1996**; 68: 1-8.
- 60) Jemal M; High-throughput quantitative bioanalysis by LC/MS/MS; *Biomed. Chromatogr.* **2000**; 14: 422-429.
- 61) Kirsch S, Bindila L; Nano-LC and HPLC-chip-ESI-MS: an emerging technique for glycobioanalysis; *Bioanalysis* **2009**; 1: 1307-1327.
- 62) Makarov A; Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis; *Anal. Chem.* **2000**; 72: 1156-1162.

- 63) Dienes T, Pastor SJ, Schürch S, Scott JR, Yao J, Cui S, Wilkins CL; Fourier transform mass spectrometry – advancing years (1992-mid. 1996); *Mass Spectrom. Rev.* **1996**; 15: 163-211.
- 64) Hu Q, Noll RJ, Li H, Makarov A, Hardman M, Cooks RG; The Orbitrap: a new mass spectrometer; *J. Mass Spectrom.* **2005**; 40: 430-443.
- 65) Madeira PJA, Florêncio MH; Applications of Tandem Mass Spectrometry: From Structural Analysis to Fundamental Studies In: *Tandem Mass Spectrometry – Applications and Principles*; InTech, published online, **2012**; 3-32
- 66) Sleno L, Volmer DA; Ion activation methods for tandem mass spectrometry; *J. Mass Spectrom.* **2004**; 39: 1091-1112.
- 67) de Hoffmann E, Stroobant V; Tandem Mass Spectrometry In: *Mass Spectrometry: Principles and Applications*; VCH, Chichester, **2007**; 189-216.
- 68) Steen H, Mann M; The ABC's (and XYZ's) of Peptide Sequencing; *Nat. Rev. Mol. Cell Biol.* **2004**; 5: 699-711.
- 69) Fabris D, Kelly M, Murphy C, Wu Z, Fenselau C; High-Energy Collision-Induced Dissociation of Multiply Charged Polypeptides Produced by Electrospray; *J. Am. Soc. Mass Spectrom.* **1993**; 4: 652-661.
- 70) Nikolaev EN, Somogyi Á, Smith DL, Gu C, Wysocki VH, Martin CD, Samuelson GL; Implementation of low-energy surface-induced dissociation (eV SID) and high-energy collision-induced dissociation (keV CID) in a linear sector-TOF hybrid tandem mass spectrometer; *Int. J. Mass. Spectrom.* **2001**; 212: 535-551.
- 71) Olsen JV, Macek B, Lange O, Makarov A, Horning S, Mann M; Higher-energy C-trap dissociation for peptide modification analysis; *Nat. Methods* **2007**; 4: 709-712.
- 72) Shao C, Zhang Y, Sun W; Statistical characterization of HCD fragmentation patterns of tryptic peptides on an LTQ Orbitrap Velos mass spectrometer; *J. Proteomics* **2014**; 109: 26-37.
- 73) An HJ, Lebrilla CB; Structure elucidation of native N- and O-linked glycans by tandem mass spectrometry; *Mass Spectrom Rev.* **2011**; 30: 560-578.
- 74) Domon B, Costello C; A Systematic Nomenclature for Carbohydrate Fragmentations in FAB-MS/MS Spectra of Glycoconjugates; *Glycoconjugate J.* **1988**; 5: 397-409.

- 75) Zubarev RA, Kelleher NL, McLafferty FW; Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process; *J. Am. Chem. Soc.* **1998**; 120: 3265-3266.
- 76) Mann M, Jensen ON; Proteomic analysis of post-translational modifications; *Nat. Biotechnol.* **2003**; 21: 255-261.
- 77) Syka JEP, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF; Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry; *PNAS* **2004**; 101: 9528-9533.
- 78) Mikesch LM, Ueberheide B, Chi A, Coon JJ, Syka JEP, Shabanowitz J, Hunt DF; The Utility of ETD Mass Spectrometry in Proteomic Analysis; *Biochim. Biophys. Acta.* **2006**; 1764: 1811-1822.
- 79) Tian Y, Zhang H; Glycoproteomics and clinical applications; *Proteomics Clin. Appl.* **2010**; 4: 124-132.
- 80) Drake RR, Schwegler EE, Malik G, Diaz J, Block T, Mehta A, Semmes OJ; Lectin capture strategies combined with mass spectrometry for the discovery of serum glycoprotein biomarkers; *Mol. Cell. Proteomics* **2006**; 5: 1957-1967.
- 81) Peracaula R, Barrabes S, Sarrats A, Rudd PM, de Liorens R; Altered glycosylation in tumours focused to cancer diagnosis; *Dis. Markers* **2008**; 25: 207-218.
- 82) Wu SW, Pu TH, Viner R, Khoo KH; Novel LC-MS² Product Dependent Parallel Data Acquisition Function and Data Analysis Workflow for Sequencing and Identification of Intact Glycopeptides; *Anal. Chem.* **2014**; 86: 5478-5486.
- 83) Mayampurath A, Yu CY, Song E, Balan J, Mechref Y, Tang H; Computational Framework for Identification of Intact Glycopeptides in Complex Samples; *Anal. Chem.* **2014**; 86: 453-463.
- 84) Catalina MI, Koeleman AM, Deelder AM, Wührer M; Electron transfer dissociation of N-glycopeptides: loss of the entire N-glycosylated asparagine side chain; *Rapid Commun. Mass Spectrom.* **2007**; 21: 1053-1061.
- 85) Baenziger NL, Brodie GN, Majerus PW; A Thrombin-Sensitive Protein of Human Platelet Membranes; *PNAS* **1971**; 68: 240-249.
- 86) Krishna SM, Golledge J; The role of thrombospondin-1 in cardiovascular health and pathology; *Int. J. Cardiol.* **2013**; 168: 692-706.
- 87) Esemuede N, Lee T, Pierre-Paul D, Sumpio BE, Gahtan V; The Role of Thrombospondin-1 in Human Disease; *J. Surg. Res.* **2004**; 122: 135-142.

- 88) Kessner D, Chambers M, Burke R, Agus D, Mallick P; ProteoWizard: open source software for rapid proteomics tools development; *Bioinformatics* **2008**; 24: 2534-2536.
- 89) Vaudel M, Barsnes H, Berven FS, Sickmann A, Martens L; SearchGUI: An open-source graphical user interface for simultaneous OMSSA and X!Tandem searches; *Proteomics* **2011**; 11: 996-999.
- 90) Cooper CA, Gasteiger E, Packer NH; GlycoMod – A software tool for determining glycosylation compositions from mass spectrometric data; *Proteomics* **2001**; 1: 340-349.
- 91) Malykh YN, Schauer R, Shaw L; N-Glycolylneuraminic acid in human tumours; *Biochimie* **2001**; 83: 623-634.
- 92) Amon R, Reuven EM, Ben-Arye SL, Padler-Karavani V; Glycans in immune recognition and response; *Carbohydr. Res.* **2014**; 389: 115-122.

VII. ABSTRACT

Glycosylation is one of the most common post-translational modifications of proteins where complex sugar chains are enzymatically attached to selected amino acid residues. Via these attached carbohydrate moieties, glycoproteins are involved in several different biological processes such as intercellular communication and immune response. Importantly, specific changes in glycosylation pattern have been observed being associated with various diseases having in this way a potential role as biomarkers. It is thus considered that analysis and understanding of this relationship between glycan structures and cell status will have some useful potential for disease diagnosis and treatment prediction.

Analysis of glycoproteins in complex samples is still remaining a challenging field in analytical chemistry. Complex structures of the glycans and its microheterogeneity within a certain protein make it necessary to get information of both, the oligosaccharide as well as the peptide backbone. This necessity, together with the absence of adequate evaluation software, and the lower ionization yield attained for glycosylated peptides, is responsible for the fact that glycoproteomics research is still far behind proteomics studies.

To tackle these problems we developed various mass spectrometric methods, including different fragmentation methods, in combination with a HPLC separation system. Fragmentation methods like CID, HCD, and ETD allowed us to get knowledge of both, glycan structure and peptide backbone. We used CID fragmentation under low energy conditions, which mainly delivers B- and Y-fragment ions of the carbohydrate moiety, for establishing the *N*-glycan structure. From these data, the mass of the peptide backbone can be calculated as well. CID fragmentation involving enhanced energy (HCD) and ETD fragmentation are then used for analyzing the amino acid sequence of the peptide backbone. Site specific glycosylation analysis on the glycopeptide level following the common bottom-up approach established in proteomics and carried out by RP-HPLC-ESI-MS/MS are shown to be a sensitive and reliable strategy for investigations on this field of research.

VIII. ZUSAMMENFASSUNG (Abstract German)

Glykosylierung bezeichnet eine post-translationale Modifikation von Proteinen, bei der Oligosaccharide enzymatisch an bestimmte Aminosäuren gebunden sind. Dadurch haben Glykoproteine Einfluss auf zellulärer Ebene wie beispielsweise in interzellulärer Kommunikation oder in immunologischen Vorgängen. Veränderungen der Glykane wurden mit verschiedenen Krankheiten in Verbindung gebracht, weshalb bestimmten Glykosylierungen eine Rolle als Biomarker zukommt. Ein funktionales Verständnis solcher Strukturen scheint daher als unverzichtbar, um Krankheiten diagnostizieren und Behandlungsstrategien entwickeln zu können.

Eine Analyse der Glykoproteine in komplexen Proteingemischen stellt nach wie vor eine Herausforderung an die analytische Chemie. Die Komplexität der Glykoproteine erfordert sowohl eine Aufklärung der Glykane als auch der Aminosäuresequenz. Solch aufwändige Analysemethoden sind zusammen mit dem Fehlen einer adäquaten Analysesoftware der Grund dafür, dass die Forschung im Bereich Glycoproteomics noch weit hinter jener der Proteomics liegt.

Für die Aufklärung der Glykoproteine haben wir verschiedene massenspektrometrische Methoden in Kombination mit einer vorgeschalteten HPLC Trennung entwickelt. Dabei kamen unterschiedliche Fragmentierungsmethoden wie beispielsweise CID, HCD und ETD zur Anwendung, um Strukturinformationen sowohl von den Glykanen als auch von den Peptiden zu erhalten. CID Fragmentierung von Glykopeptiden liefert fast ausschließlich B- und Y-Fragmente der Oligosaccharide, die eine Aufklärung der N-Glykane erlauben. HCD und ETD Fragmentierungen wurden für die Analyse der Aminosäuresequenz herangezogen. Die Aufklärung von Glykosylierungsmuster auf Glykopeptidebene hat gezeigt, dass sich die entwickelten RP-HPLC-ESI-MS/MS Methoden für weitere Untersuchungen auf diesem Forschungsgebiet eignen.

IX. CURRICULUM VITAE

Personal data

Name: Dietmar Hammerschmid
Date of birth: 04/06/88
Place of birth: Linz (AUT)
Citizenship: Austria

88

Education

03/2014 – now **Master Thesis**
Department of Analytical Chemistry
Group of Prof. Rizzi
"Targeted glycoproteomics for monitoring changes in glycosylation upon cell activation"

03/2012 - now **Masters Programme in Biological Chemistry**
University of Vienna

10/2007 – 02/2012 **Undergraduate Studies in Nutritional Science**
University of Vienna
Bachelor Thesis: "Genetic influences on obesity"

09/2002 – 06/2007 **Technical High School**
HTBLA Perg (AUT)
focus on: Software Engineering and Organisation

Languages

German Native language
English fluent in speech and writing
