universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Impact of OST catalytic isoforms STT3A and STT3B on global glycosylation and ER homeostasis“

verfasst von / submitted by

David M. Hoi, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2018 / Vienna 2018

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 863
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Biologische Chemie
Betreut von / Supervisor:	Univ.-Prof. Dr. Christopher Gerner

Acknowledgements

I want to thank Karl Mechtler for giving me the opportunity to do my master thesis in his group at the IMP and offering me a unique working field with every possibility a scientist could ask for.

I also would like to thank Jasmin Taubenschmidt for her support in cell culture and her biological input as well as Joseph Penninger for the possibility to work with his team and lab.
Thanks to my colleagues and now friends, that helped me a lot during this period and made longer working days exciting and less stressful.
Furthermore, I want to thank Veronika, Maria, Julia and Elfi, who supported me during this year, although I had too little time for them.

Finally, but most importantly, I would like to thank Johannes Stadlmann, the most amazing supervisor anyone could wish for. He was there 24/7 to support me and my work, to answer my questions and to guide me in the right direction, so that I would get good results and the best possible outcome for my time at the institute.

Abstract

Eukaryotic glycosylation machinery has been extensively studied in the past. Combining structural information on the complex assembly and using biochemical assays to follow glycosylation behavior of the OST targets have led to the identification of two distinct catalytic subunits. STT3A has been found to act co-translationally and to be responsible for most glycosylation events in the cell. STT3B, however, should post-translationally glycosylate skipped sites or sequons, that are not easily accessible. However, proteomics studies, that cover the general glycosylation development upon depletion of either one of the two catalytic subunits, are still missing.

On a mouse embryonic stem cell knockout model for both STT3A and STT3B, a comparative and quantitative glycoproteomics workflow was developed to cover the glycosylation response at intact glycopeptide level. This approach involves a hpH-chromatographic separation of peptides and a HILIC separation for glycopeptide enrichment. Data analysis of intact glycopeptides was performed with the SugarQb software.

STT3B dependencies of proposed sites could not be determined with this approach using this mESC model. Still, a response in glycosylation could be observed at protein level. Both knockout models where viable and capable of maintaining glycosylation levels by compensating it due to increase in the each other subunit. Although certain proteins, such as many ER residential proteins of the ERAD machinery, show a dependency of certain glycosylation subunits, STT3B is still able to outcome its role as attendant modifier.

Zusammenfassung

Der Glycosylierungs-Mechanismus in Eukayoten wurde in der Vergangenheit intensiv erforscht. Anhand von strukturellen Daten über den Oligosaccharyltransferase Komplex und dessen Zusammensetzung, sowie über biochemische Assays zur Studie der Glycosylierung, wurden zwei verschiedene katalytische Untereinheiten entdeckt. Die STT3A Untereinheit wird als co-translational beschrieben, während STT3B post-translational glycosyliert, wenn bestimmte Stellen am Protein von STT3A ausgelassen werden. Proteomische Studien, die das Glycosylierungsverhalten der beiden Untereinheiten während des Knockouts der jeweils anderen beschreiben, fehlen jedoch.
Für ein Knockout Modell in mESCs für STT3A and STT3B wurde ein Glycoproteomischer Workflow entwickelt, um intakte Glycopeptide quantitativ zu messen. Eine hpH Chromatographie wurde verwendet um Identifikation durch optimale Separation von Peptiden zu erhöhen. Als Anreicherungsmethode für Glycopeptide wurde ein HILIC System verwendet. Die Datenanalyse wurde mit der SugarQb Software durchgeführt, welche die Identifikation von intakten Glycopeptiden erlaubt.
Die zahlreichen Spezifikationen die für STT3B in früheren Studien gefunden wurden, konnten in diesem Modell nicht nachgewiesen werden. Allerdings konnte eine Änderung auf Glycosylierungsebene sehrwohl festgestellt werden. Der Knockout beider Untereinheiten produzierte Zellen mit normaler Überlebensrate und die Glycosylierung konnte durch Kompensation der jeweils anderen Untereinheit aufrechterhalten werden. Insbesondere ER-resident Proteine und jene des ERAD, wiesen Abhängigkeiten für beide katalytischen Untereinheiten des OST Komplexes auf. Die hohe Aktivität von STT3B im STT3A Knockout zeigt, dass STT3B dazu in der Lage ist, mehr als nur bestimmte Positionen im Protein zu glycosylieren, die von STT3A übersprungen werden.

Contents

List of Abbreviations

CAM	Cell adhesion molecule
CDG	Congenital disorder of glycoslyation
CID	Collision induced dissociation
ER	Endoplasmic reticulum
ERAD	ER associated
ESI	Electrospray ionization
ETD	Electron transfer dissociation
FDR	False discovery rate
HCD	Higher-energy collisional dissociation
HILIC	Hydrophilic interaction chromatography
iTRAQ	Isotope tag for relative and absolute quantification
mESC	Mouse embryonic stem cells
MRM	Multiple reaction monitoring
MS	Mass spectrometry
MS/MS	Tandem MS
OST	Oligosaccharyltransferase
PRM	Parallel reaction monitoring
PSM	Peptide spectrum match
PTM	Post translational modification
SILAC	Stable isotope labelling by amino acids in cell culture
TMT	Tandem mass tag
UPR	Unfolded protein response

1 Introduction

1.1 Introduction to glycobiology

Protein glycosylation, among other PTMs, such as phosphorylation, acetylation or methylation, is considered as one of the most diverse protein modifications. It provides heterogeneity not only towards its binding residues, but also within its macromolecular structures. Numerous possible assemblies of different sugars attached to proteins are possible by linking the oligosaccharides to the amino acid asparagine via its sidechain nitrogen (N-linked glycosylation) or threonine and serine via its sidechain oxygen (O-linked glycosylation). That causes a variety of biological functions in the cell, which can differ between organisms and tissues. In mammalian cells, glycan structures comprise 10 different types of monosaccharide, which are glucose (Glc), galactose (Gal), mannose (Man), fucose (Fuc), N-acetylglucosamine (GlcNAc), N-acetylgalactosamine (GalNAc), xylose (Xyl), glucuronic acid (GlcA), iduronic acid (IdoA), and 5-N-acetylneuraminic acid (Neu5Ac) [1]. Glycans play a substantial structural and modulatory role when it comes to protein secretion and folding. This is primarily achieved during protein quality control in the endoplasmic reticulum (ER), by providing ligands for glycan binding chaperons. Moreover, cell adhesion is often driven by glycan anchoring, either necessary for cell-cell recognitions and interactions, or for cell-matrix adhesion, which is mediated by glycan binding by cell adhesion molecules (CAMs) such as selectins, integrins or cadherins [2]. These interactions can be both intrinsic and extrinsic, which is important to differentiate between pathogenic microbes and intracellular glycan patterns during pathogen invasion [3]. However, glycoproteins make up a large part of membrane proteins and often, receptor associated glycans directly alter protein function and activate signaling cascades, that regulate cytosolic and nuclear functions [2]. Moreover, glycans are strongly associated with the vertebrate immune system and are involved in immuno-regulatory functions. Selectin mediated leukocyte trafficking as well as TCR/BCR signaling, antibody functionality, auto-immunity and immune-cell differentiation are highly regulated by the interplay of glycan epitope presentation and their recognition by glycan binding proteins (GBPs) [4, 5].

Consequently, glycosylation as PTM is connected to many physiopathological processes, linking altered glycosylation to various kind of disease and disorders, such as comprised immune system and inflammatory reactions, developmental abnormalities or cancer. Cancer initiation and progression involves distinct molecular and cell biologic processes. Many of them demand on glycosylation mediated regulation, like cell signaling, tumor cell dissociation and invasion, cell-matrix adhesion, angiogenesis, immune response and metastasis. Since the heterogeneity is derived from different aberrations in glycan composition and protein-, site- and cell type specificity, a wide range of glycosylation alterations can be defined in different cancers. Carbohydrate alterations in cancer are typically caused by incomplete glycan synthesis, leading to truncated glycan structures or carbo-hydrate neo-synthesis. Altered induction of carbohydrate presentation related genes, such as glycosyltransferases and glycosidases is often a consequence [1, 6].

Congenital disorder of glycosylation (CDG) covers a set of diseases, which are primarily associated with hypoglycosylation and in most cases the N-glycosylation pathway is affected. However,

several subtypes also affect O-glycosylation or a combination of O- and N-glycosylation pathways. CDGs often involve developmental deficiencies, comprised brain developmental, psychomotor retardation and endocrine abnormalities [3]. Two types of CDG syndromes can be distinguished that differ in their specific alteration during the processing steps. Type I CDGs are mainly associated with assembly defects. Proteins often appear to be hypoglycosylated, missing the entire carbohydrate chain due to deficiencies during the many steps in synthesis of lipid-linked oligosaccharides in the ER. Mutations interfering with oligosaccharide synthesis are commonly monoallelic, since a complete knockout of key genes involved in N-glycosylation is lethal. Type II CDG disorders are less consistent because they appear as glycan processing defects. Mutations of different targets during oligosaccharide processing in the cell by glycosyltransferases, sugar transporters and cytoplasmic proteins can cause CDG-II disorders [7–10]. Type I related defects in oligosaccharyltransferase, which is necessary for oligosaccharide transfer onto proteins entering the ER, have been found in four different subunits of the oligomeric complex (TUSC3, MAGT1, DDOST, STT3). Moreover, mutations in DDOST and STT3A are known to result in glycosylation abnormalities of transferrin

1.2 Protein N-glycosylation

N-glycosylation biosynthesis pathway in eukaryotes is performed in two compartments, the endoplasmic reticulum [Fig. 1] and the Golgi. In a first step, the precursor synthesis, Dol-P is formed by flipping Dol-P-P (consisting of 19 isoprene units in eukaryotes) to the cytoplasmic side of the membrane and dephosphorylating it. In a series of glycosylation events, precursors get modified each by a single sugar entity, transferring it from nucleotide sugars UDP-GlucNac and GDP-Mannose, a process catalyzed by enzymes of the ALG (Asn-linked glycosylation) class. The synthesis is carried out in the cytoplasm until a $Man_5GlcNAc_2$-P-P-Dol structure is formed. After that, the structure gets translocated to ER lumen by the RFT1 flippase. Further extension of the sugar structure by four mannose- and three glucose residues is achieved by different ALG enzymes. Precursor is provided by glycosylation of dolichol at cytoplasmic side and translocated to the lumen with support of MPDU1. Members of the ALG enzyme group are responsible for the synthesis progression, forming firstly a $Man_9GlcNAc_2$-P-P-Dol and secondly a $Glc_3Man_9GlcNAc_2$-P-P-Dol. Transfer of sugar residues onto potential N-glycosylation acceptor sites (Asn-X-Ser/Thr/Cys) is driven by oligosaccharyltransferase complex (OST) [3].

Glycoproteins undergo protein folding by passing the Calnexin/Calreticulin-cycle in the ER. Here, the last three glucose residues get removed and can than continue to Golgi for further processing. Most glycoproteins entering the cis-Golgi carry glycan structures containing $Man_9GlcNAc_2$ or $Man_8GlcNAc_2$. Glycan trimming to $Man_5GlcNAc_2$, an important player in the generation of complex and hybrid glycan structures [Fig. 2a], is achieved by α-mannosidase I (α1-2 sensitive). Biosynthesis of such structures is then performed in the medial-Golgi. Glucose residues are sequentially added by N-acetylglucosaminyltransferase (MGAT1), whereas α-mannosidase II with α1-3 and α1-6 sensitivity (MAN1A1 and MAN1A2) are removed. Complex structures can be built by adding sugars like fucose, sialic acid, galactose and N-acetylglucosamine or phosphate and sulfate on the Man5GlcNAc2-core

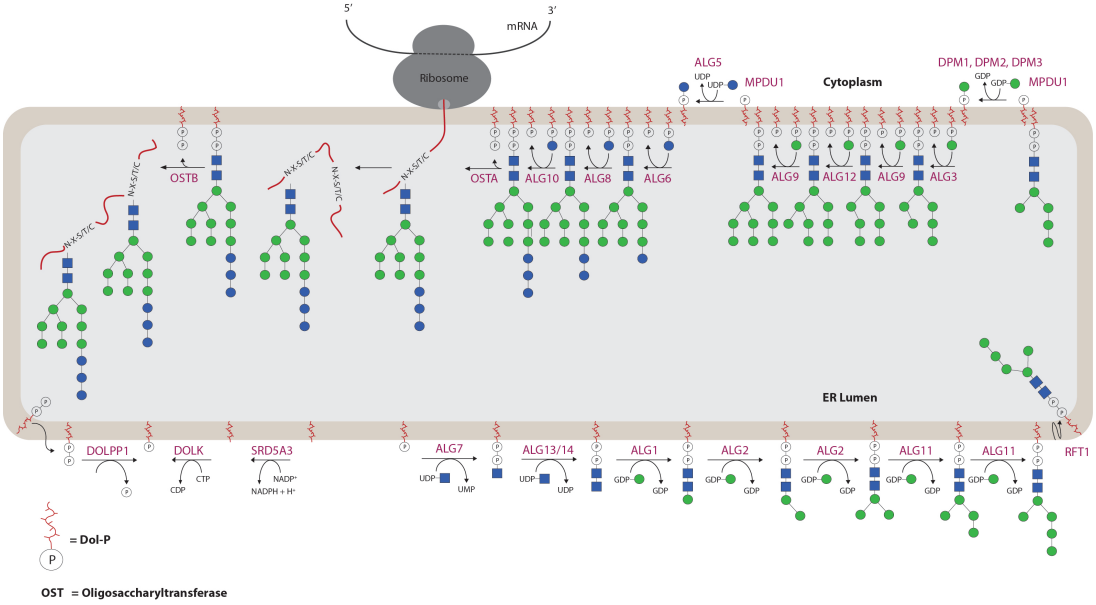


Figure 1: N-glycosylation pathway in eukaryotic systems (modified from [3])

structure. However, hybrid structures are formed due to incomplete removal of mannose residues from glycans.

SEC61 complex (forming the main ion-channel), together with the translocon associated protein complex (TRAP), TRAM and OST, forms the translocon complex. With TRAP and the OST subunit RPN1 the translocon couples to ribosomes from which nascent polypeptide chains are transported into the ER. In mammalian cells two isoforms of the membrane associated oligosaccharyl transferase have been determined, differing in function and cellular localization. Along with a shared group of subunits (OST4, TMEM258, OST48, RPN1, RPN2 and DAD1) they consist of different catalytic subunits (STT3A and STT3B). In addition, several isoform specific subunits exist, with functions that have not been described in completion (DC2 and KCP2 - STT3A; TUSC3/MagT1 - STT3B). OST isoform that contains STT3A is known to be necessary for bulk glycosylation. Modification at the according motif (NxS/T/C) occurs co-translationally, during the moment the polypeptide chain enters the ER lumen through the SEC61 channel. Protein glycosylation sites, that are skipped by STT3A can then be modified by STT3B. Cryo-electron microscopy based structural analysis studies [Fig. 2b] of the whole translocon complex as well as the OST complex alone, suggest translocon association only for OST isoform containing STT3A [11]. This finding supports the hypothesis of STT3B being responsible for post-translational glycosylation, for residues, that had been skipped by STT3A. Post-translational specificity of STT3B has been shown in several studies and point towards selective terminal glycosylation. Glycosylation sites, that are located closer than 75 amino acids to the C-terminus apparently can be skipped by STT3A. OST activity is achieved, when 65-75 amino acids are between the glycosylation acceptor site and the peptidyl-transferase of the large ribosomal subunit. Site skipping is thought to be promoted by a more rapid transfer of the terminated peptide chain [12]. N-glycosylation sites, that are located close to the signal peptide sequence of secretory

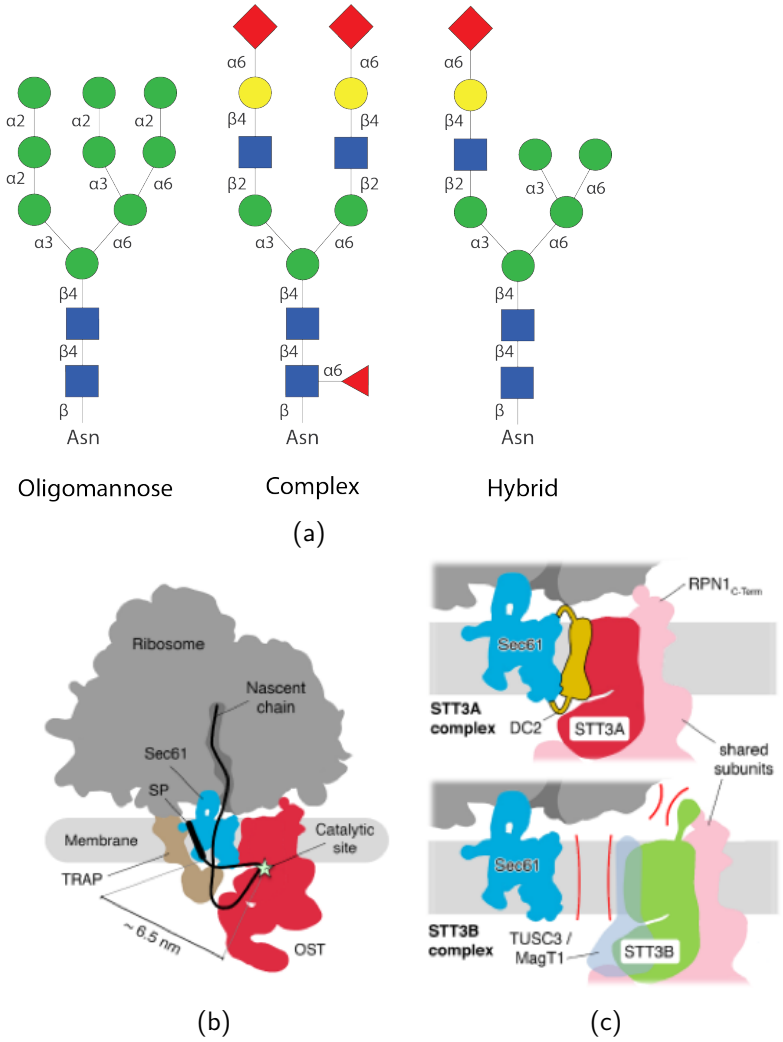


Figure 2: a) Possible oligosaccharide assemblies for N-glycosylation in eukaryotes. b) Translocon complex including ribosome, OST complex and TRAP. c) Visualization of the post-translational and co-translational activity of STT3A and STT3B [11].

proteins have been found to be dependent of STT3B glycosylation as well [13]. A bivalent glycosylation specificity can be observed on closely spaced acceptor sites. Closely spaced acceptor sites have been found to be reliant on co-translation glycosylation, when there are optimal conditions. Glycosylation efficiency is comprised when sequons are adjacent or contain serine as modified amino acid residue such as in "NxTNxT" and "NxSANxS" [14]. Moreover, the type of middle amino acid in sequons are assumed to play a role in co-translational glycosylation. Glycosylation kinetics of NxS sequons containing large hydrophobic or negatively charged amino acids (W, L, F, I, M and D) at middle position is shifted towards post translation modification. NxT containing sequons are reliant on STT3A glycosylation [15].

1.3 Protein folding and ER - associated degradation

Protein folding is an important part of protein biosynthesis and mainly takes place in the cytoplasm. Formation of secondary structures, as well as disulfide bonding and association to oligomeric complexes are crucial. Cytosolic protein folding, repair and multimerization involves chaperones from the heatshock protein families hsp60 and hsp70. Hydrophobic regions on proteins are bound by the chaperones, allowing better solubility to achieve correct folding. However, membrane and secretory proteins that often carry glycan structures, can be translocated post translationally or co-translationally via the translocon from membrane associated ribosomes into the ER lumen. Protein folding is then promoted by a molecular machinery of chaperones (Grp78, Grp94 and Grp170) and protein disulfide isomerases (PDI, ERp57, ERp59 and ERp72) [3].

During ER folding processes N-glycosylation plays the important role of being recognized by the glycan binding domain of the calcium dependent lectin-type chaperons calnexin (CNX) and calreticulin (CRT) [Fig. 3]. Whereas calnexin is membrane associated and is crucial for retaining immature membrane located glycoproteins, calreticulin functions on proteins in the ER lumen. Monoglucosylated oligomannose structures on proteins are detected and bound by these lectins. Protein folding of CNX/CRT-captured proteins is supported by a network of chaperones and protein disulfide isomerases like BiP/Grp78, Grp94 and ERp57. Removal from CNX of CRT is achieved during cleaving of the terminal glucose by a glucosidase 2 (GII) [3]. Folded proteins can then either be exported to Golgi or in case of misfolding be detected by UDP-glucose glycoprotein glucosyltransferase 1 (UGGT1). UGGT selectively targets hydrophobic surfaces, thus recognizes misfolded proteins and reattaches the terminal glucose residue. Proteins carrying such a sugar motif can then reenter the CNX/CRT cycle for proper folding. Kinetic competition of re- and de-glucosylation with ER-mannosidase 1 (ER-ManI) enables terminally misfolded proteins to exit this cycle [16]. It further facilitates EDEM1 and EDEM3 binding, resulting in oligosaccharide structures which are incompatible with UGGT binding. EDEMs are lectins with $Man_8GlcNAc_2$ specificity. BIP targets the hydrophobic surface of misfolded proteins, together with other chaperones and PDIs. Trimmed oligosaccharides are recognized by the mannose-6-phosphate receptor homology domain (MRH) of the ER residential lectins OS-9 and XTP3-B [17,18]. OS-9 and XTP3-B have been found to bind the ER membrane glycoprotein SEL1L, linking them to Hrd1, the main component of the dislocon together with a E3 RING finger ubiquitin ligase domain. VCP/p97 complex is necessary for final substrate dislocation, after which the ubiquitinylated proteins get degraded by the 26S proteasome [18].

1.4 Glycoprotein analysis

To evaluate the glycosylation state of a proteome, with focus on N-glycosylation as PTM, bioanalytical methods have been driven towards both the identification of specific sites and the glycosylation form. Mass spectrometry based proteomics has evolved as the primarily used strategy for answering fundamental questions in glycobiology. Glycosylation levels are used to refer to a specific functional state of a cell. They help to understand protein synthesis, processing and their distribution throughout the cell. With such information, links can be drawn from abnormalities at genomic level to

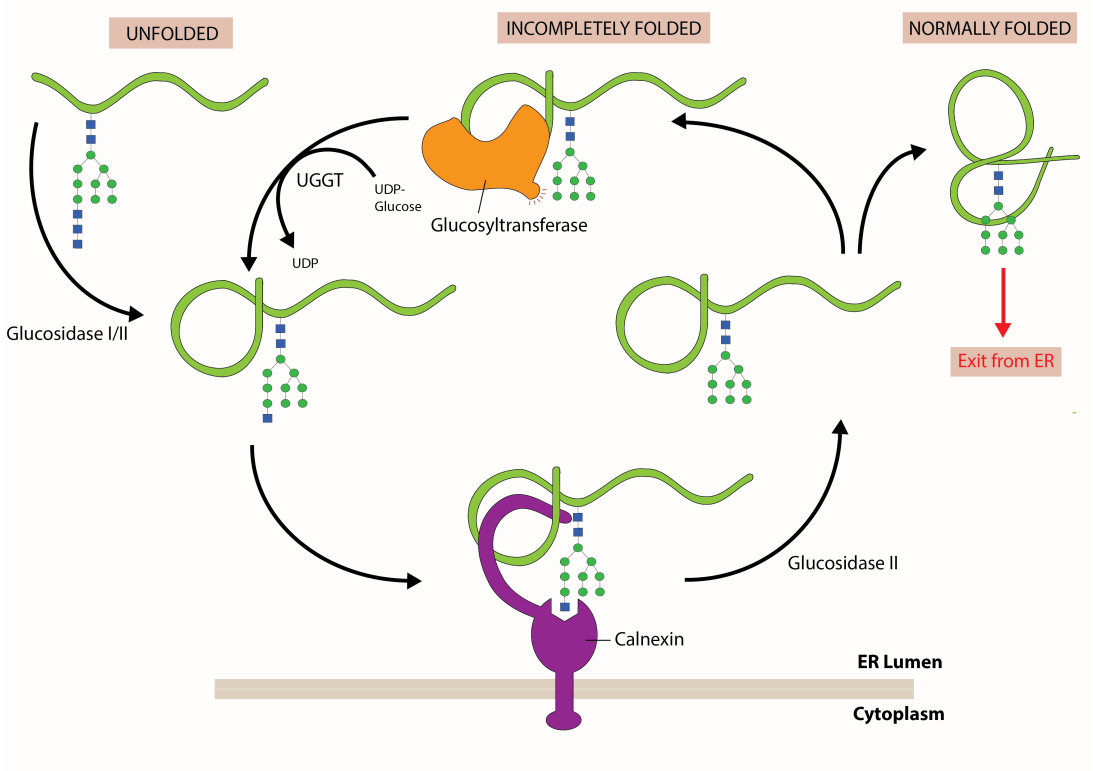


Figure 3: Calnexin/reticulin cycle for protein folding of glycoproteins in the ER.

protein irregularities, that are reflected by altered glycosylation behavior. However, analytical technology in glycobiology should not only be used for global approaches, but also for detection of certain glycosylation markers, that can be assigned to types of disease.

1.4.1 Glycopeptide enrichment strategies

Prior to proteomics experiments, in most cases it is necessary to decomplex the sample. Taking into consideration their manifold appearance, enrichment of glycosylated peptides requires a series of analytical strategies, that have been adapted for a diverse set of challenges. Glycoproteomic downstream processing can vary in sample preparation in aspect to sample material, enrichment and bioinformatic analysis according to the scientific question.

Cellular localization of glycoproteins as well as their localization in vivo makes them less prone for conventional tissue extraction and cell lysis methods. In vivo, glycoproteins as well as proteoglycans are often found in mucus rich tissues like lung, colon and pancreas. Therefore, methods are necessary, that are capable of disrupting mucine secreting mucosa and efficiently extracting glycoproteins by separating them from mucines. In general, samples are treated fresh, but also FFPE-tissue is widely used in the proteomics field. Different extraction methods have been developed, that can be used both for paraffine fixed and fresh samples. They are mainly relying on extraction with organic solvent like chloroform/methanol [19] or xylene/methanol [20]. Rehydration of extracted sample is mainly done using ethanol with an increasing aqueous phase.

A large proportion of glycosylated proteins appears to be membrane associated, thus detergent based extraction is often used to increase membrane proteins in general. Due to interference, cell lysates containing high amounts of SDS or other detergents are not suitable for mass spectrometric analysis [21]. Proteomic studies can overcome such problems, with stringent washing procedures using C18 cartridges [22]. In glycoproteomics, low concentration and abundance in a large dynamic range often is a limiting factor, therefore sample losses are meant to be held small. A method, allowing to completely remove detergents is FASP. This method uses ultrafiltration membranes to remove salt, detergents and every other low MW molecules [23]. FASP combines lysis and washing steps with protein digestion, which is also done on the filter after alkylation and reduction. The group of Matthias Mann also developed a FASP protocol which is suitable for glycoproteomic analysis. Original FASP is extended by an affinity step, where lectins are used to capture digested glycopeptides [23]. PNGase F is used to cleave glycans of the peptides, which can then be eluted from the column. Glycans and undigested material like DNA get retarded by the column.

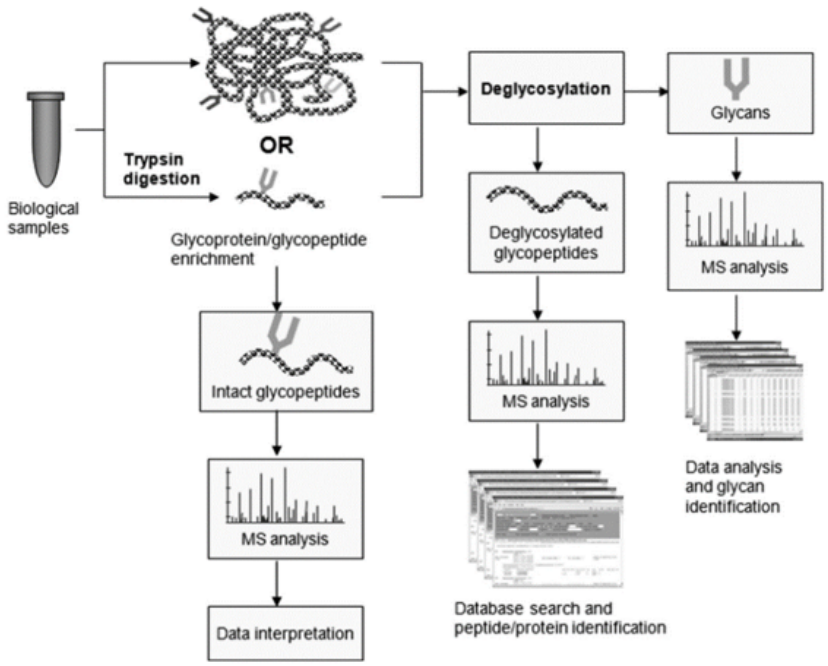


Figure 4: Traditional workflow, that uses either intact glycopeptide identification or separation of glycans from the peptides [24].

Information that can be generated by PNGase F based methods like FASP are restricted to glycosylation site identification, whereas knowledge on glycan structures gets lost [Fig. 4]. Independent of the glycopeptides processing, both deglycosylation based and intact approaches are generally bottom-up proteomics strategies. A cell lysate or extracted tissue is digested with the endoprotease trypsin, which has a C-terminal cleavage specificity for arginine and lysine residues. Prior to digestion, proteins containing cysteine-cysteine disulfide bonds get reduced by DTT and the free cysteines get alkylated with IAA. For studies, aiming for glycosylation site identification, PNGase F, a peptide-N4-

(N-acetyl-beta-glucosaminyl) asparagine amidase, is used to catalyze the glycosidic bond cleavage between the asparagine and the core GlcNAc of the attached sugar. During glycan removal deamidation of asparagine to aspartic acid occurs. One solution to distinguish between enzymatically induced and naturally appearing deamidation is the usage of 18O-water during the deglycosylation step. Hereby heavy oxygen will be incorporated into every newly built aspartic acid [23]. In addition to PNGase digest, a glycomics approach can be performed, giving insight on glycan composition of the sample, but no assignment can be made.

Although the identification of intact glycopeptides remains challenging in bioanalytical aspects, several purification strategies have evolved, which allow for selective enrichment of glycosylated peptides and depletion of the non-glycosylated excess.

1.4.2 Lectin affinity chromatography

Lectins, a large group of carbohydrate/binding proteins, have been shown to possess affinities for various oligosaccharide epitopes. About 160 lectins have been identified mainly from plants and for many of them the specificity is known, although selectivity does not always seem to be exclusive [25, 26]. The chromatographic application of these distinct and very sensitive sugar-protein interactions allows the separation of different kinds of glycosylated analytes. A large range of lectins, that are also commercially available, is currently used for chromatographic purpose [26]. Concanavalin A (Con A) has a binding affinity for mannosyl and glucosyl containing sugar residues, whereas wheat germ agglutinin (WGA) targets N-acetyl-glucosamine and sialic acid and jacalin (JAC) specifically binds to galactosyl (-1,3) N-acetyl-galactosamine [Table 1].

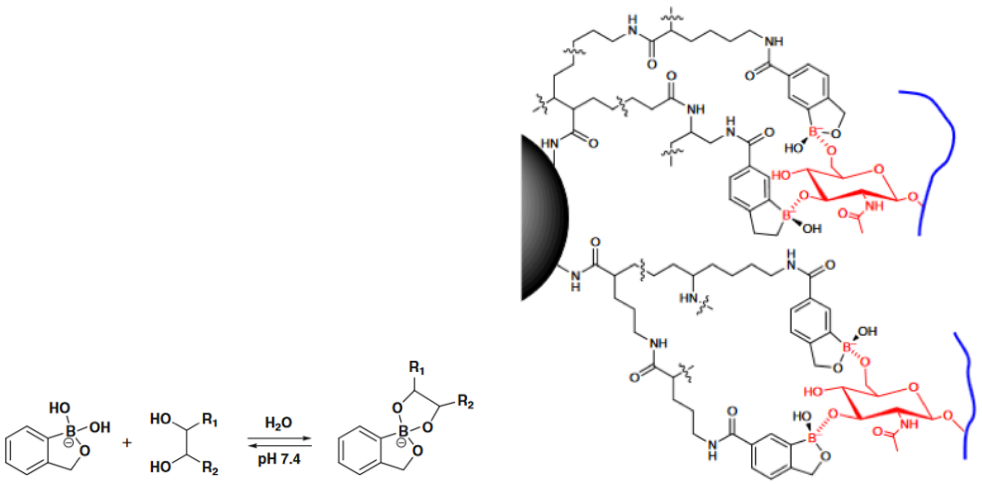
Lectin type	Name	Source	Affinity
Mannose binding lectins	Con A (Concanavalin A)	Canavalia ensiformis	High-mannose, hybrid and biantennary complex type N-glycans
Fucose binding lectins	UEA (Ulex europaeus agglutinin)	Ulex europaeus	Fuc α1-2Gal-R
Galactose/ N- acetylgalactosamine binding lectins	AIL (Jacalin)	Artocarpus integrifolia	(Sia)Gal β1-3GalNAc α-1-Ser/Thr
Sialic acid/ N-acetylglucosamine binding lectins	WGA (Wheat Germ agglutinin)	Triticum vulgaris	GlcNAc β1-4GlcNAc β1-4GlcNAc, Neu5Ac (sialic acid)

Table 1: Different lectin types and their affinities. [27]

One limitation of lectin affinity chromatography is its restriction in terms of glycoproteome coverage. There have been attempts to combine different lectins, in order to target a broader spectrum of glycoforms. Such approaches are either using a serial chromatographic workflow or multiple lectins in one column, which also enables combinatory enrichment of N-glycopeptides and O-glycopeptides [28, 29]. However, elution conditions do not always meet binding specificities, allowing elution of α-fucose containing glycopeptides from some plant lectins with mannose containing buffer [25].

1.4.3 Boronic acid chemistry

A more generic and chemical approach for glycopeptide enrichment is based on the binding affinity of arylboronic acids with glycans. Ortho-substituted hydroxyalkyl arylboronic acid has been found to have a high specificity for reducing sugars like glucose in a glycan chain [30]. The binding appears due to the generation of cyclic esters between boronic acid and diol groups on the sugar residue, where both hydroxy groups are involved in the heterocycle formation. Glycopeptide capture can easily be regulated by shifting the pH to a neutral or slightly alkaline level, which puts the ortho-localized hydroxy group in an anionic state [Fig. 5a].



(a) (b)

Figure 5: a) Glycopeptide capture via cis-diols during pH shift [30]. b) 1-hydroxy-1,3-dihydrobenzoxaboroles) coupled to dendrimers of L-lysines on amino functionalized magnetic beads [31]

Boronic acid-based approaches enable enrichment of a broader spectrum of glycopeptides, since no specific recognition motif is needed except for the presence of diol groups [31]. O-glycosylated peptide enrichment has been challenging using this technique, because boronic acid interaction with sugars in the absence of cis-diols, like GlcNAc in O-glycans is much weaker. This limitation had been overcome by the usage of benzoboroxole (1-hydroxy-1,3-dihydrobenzoxaboroles), coupled to dendrimers of L-lysines on amino functionalized magnetic beads [Fig. 5b]. This approach showed a large increase in identification of O-glycopeptides as well as N-glycopeptides [32].

1.4.4 Titanium dioxide affinity chromatography

In phosphoproteomics approaches titanium dioxide (TiO2) is widely used for enrichment of negatively charged phospho-entities, that interact with the charged titan(IV)-ion. The concept of immobilized metal affinity chromatography (IMAC) has been adapted for several applications in the field of high throughput mass spectrometry, targeting acidic compounds, such as small hydrophilic compounds, phospholipids, glycosylphosphatidylinositol, phospho-derivatized O-linked N-acetylglucosamine containing peptides and sialylated glycopeptides [33, 34]. Given the heterogeneity of analytes affected

by this chromatographic procedure, there are suggestions on reevaluation of generated data, like searching for sialiome analysis in a phosphoproteome dataset [35]. Optimization of the protocol towards sialic acid enables the specific enrichment of a large population of O- and complex N-linked glycopeptides.

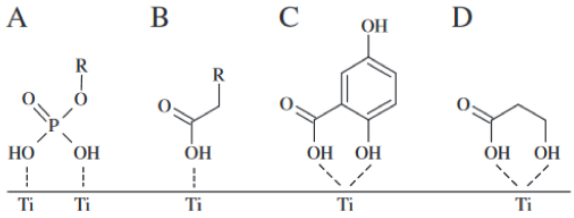


Figure 6: Coupling variants to immobilized TiO2 [34]

Either on beads or in tip-based systems, sialic acid containing glycopeptides are bound to immobilized TiO2 under acidic conditions [Fig. 6]. Successful enrichment of glycopeptides presupposes the depletion of phosphorylation sites by alkaline phosphatase. With TFA a pH in the range of sialic acid (2.60) and aspartic (2.77) as well as glutamic acid (3.22) is adjusted. To minimize unspecific binding of amino acid side chains, glycolic acid is used, which has a stronger affinity towards metal ions compared to other carboxylic acids. Elution can be achieved by shifting to higher pH with ammonia solution [36].

1.4.5 Hydrazide chemistry

Another strategy to covalently bind glycoproteins for selective enrichment is based on capturing them to a solid phase by hydrazide chemistry, originally developed in Aebersolds group for analysis at protein level [37]. Prior to binding, the glycoproteins must undergo an oxidation step with sodium periodate, converting cis-diol groups of the non-reducing sugars to aldehyde groups. Via hydrazone formation, the aldehydes get coupled to hydrazide groups that are immobilized on a solid phase, on beads or a membrane [Fig. 7].

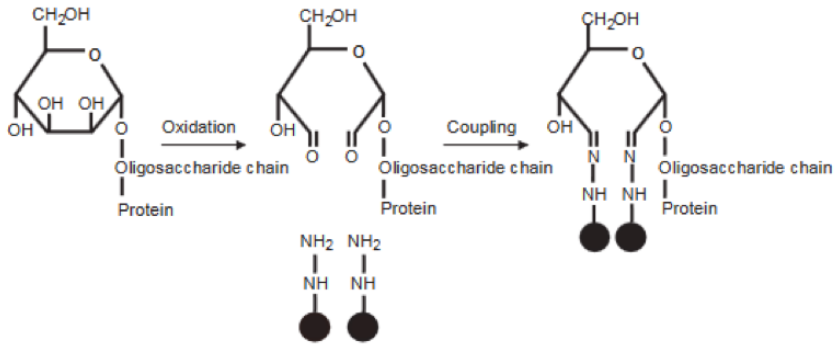


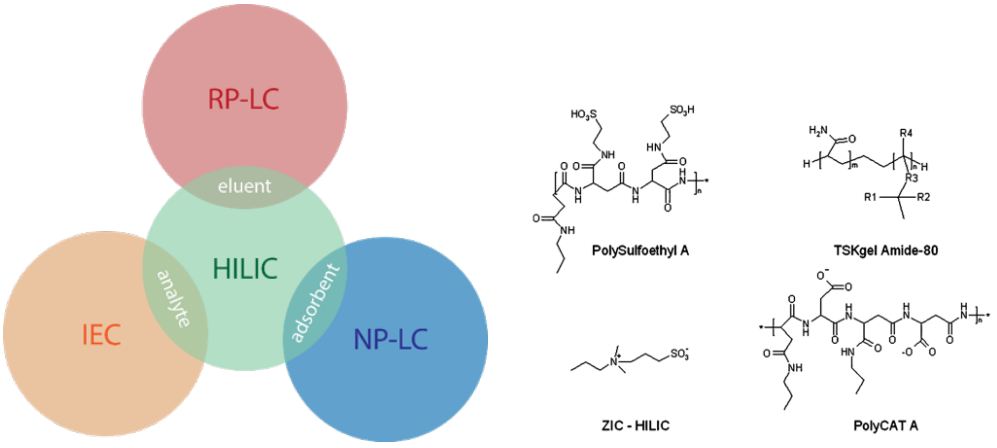
Figure 7: Binding of non-reducing sugars to beads by oxidation to aldehydes and coupling with hydrazide groups [37].

In a series of stringent washing steps, the non-glycosylated proteins get removed, leaving glycoproteins only on the solid phase to get proteolytically digested. The original approach intends isotopic labelling of the glycopeptides for quantification, where each the α-amino groups get labelled with different isotopic forms of succinic anhydride. The peptides then get released by PNGase F and can be analyzed in an LC-MS/MS experiment. In newer approaches, the selectivity for carbohydrates is used on peptide level. Glycoproteins get digested prior to the enrichment and immobilized on solid particles as in solid-phase extraction of glycopeptides (SPEG)

1.4.6 Hydrophilic interaction liquid chromatography

HILIC mechanism was originally described in 1975 by Landen et al. and named by Alpert et al. in 1990, with the intention of separating polar analytes between a hydrophilic stationary and a hydrophobic mobile phase. It offers an arrangement that reminds of a normal phase liquid chromatography (NP-LC), since it employs similar polar stationary phases like silica, amino and cyano material. Compared to NP-LC, HILIC is mainly operated in an aqueous system with water-miscible organic phase. In opposition to reversed phase liquid chromatography (RP-LC), elution is achieved by a gradient of increasing polar aqueous phase. Due to the higher solubility and good separation of polar compounds in comparison to RP-LC systems, HILIC has emerged as the main method for analyzing uncharged and highly hydrophilic analytes, that are not suitable for ion-exchange chromatography (IC), because of low charge states. Therefore, HILIC is often described as the method, combining properties of NP-LC, RP-LC and IEC [Fig. 8a] [38].

HILIC systems have been adapted for separation of challenging analytes, like polar drugs, metabolites



(a) (b)

Figure 8: a) Overlap between different chromatographic systems and their contribution to HILIC. b) Different commonly used materials for HILIC columns [39].

and carbohydrates, and the technology has been driven towards material and method development since [39]. Several stationary phases have been developed for different HILIC applications, mainly silica based, like PolySulfoethyl A and PolyCAT A (cation exchanger), PolyWAX (anion exchanger), TSKgel Amide-80 and the zwitterionic (ZIC)-HILIC [Fig. 8b] [39–42]. Although, chromatographic

behavior like retention and selectivity differs between these methods, they share a similar mobile phase, as well as the same separation principle [43]. The separation is thought to be driven by the enrichment of water in the polar stationary phase surface from the HILIC buffer. In the most cases acetonitrile (ACN) and water is used, starting with an ACN content above 70% [39].

HILIC solvent strengths [39]:
Acetone < Isopropanol < Propanol < Acetonitrile < Ethanol < Dioxane < DMF < Methanol < Water

In contrast, a water-deficient mobile phase is created, which leads to partitioning and distribution of the hydrophilic analytes between both phases in the form of liquid/liquid extraction [Fig. 9]. Increase of hydrophilicity to elute the analyte is achieved by increasing the water content in the mobile phase. Additionally, depending on the stationary phase, a combination with either changes in electrostatic interactions or hydrogen bonding between neutral polar species can be used as well [44].

Apart from the choice of the stationary and mobile phase also salt concentration and pH play

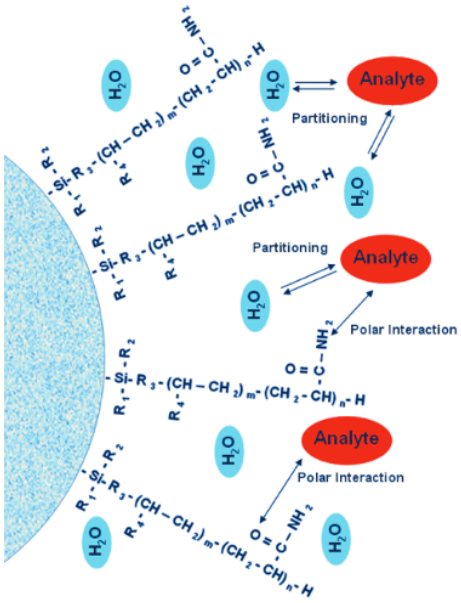


Figure 9: Tosoh TSKgel®AMIDE-80 2 µm HILIC COLUMNS; [45]

an important role in the HILIC mechanism, for which mainly MS-compatible ionic additives like ammonium acetate and ammonium formate are used. In stationary phases, that interact with the analytes by electrostatic interactions, such salt additives can be introduced into the system via mobile phase to decrease these interactions [38].

Hydrophilicity makes glycopeptides, which appear in a large heterogeneity, an optimal candidate for separation using a HILIC system [43]. Retention of glycopeptides generally depends on their polarity. It is determined by several factors, like glycan composition, charges due to sialic acid residues, and their general size and branching. However, according to the number of polar group, HILIC for

glycopeptides still is reproducible towards retention time prediction. Especially in glycan analysis and interpretation HILIC offers great potential for structural determination. Moreover, HILIC can be used in combination with quantitative analysis strategies like ITRAQ or TMT-tags, or fluorescence tags, with the side effect of shifting the analytes to a less hydrophilic state [46].

In HILIC approaches, TFA has been found to be a suitable ion-pairing reagent to improve separation efficiency [47]. Studies have shown, that charged and nonglycosylated residues, such as negatively charged peptides (C-terminus, aspartic acid and glutamic acid) and positively charged peptides (N-terminus, lysine and arginine) can increase hydrophilicity and create an overlap with the actual analyte, which leads to coelution and a comparably bad separation. Addition of low-pH ion pairing reagents like TFA help to overcome this problem, by binding to positively charged residues and by protonating negatively charged ones [Fig. 10]. Firstly, electrostatic interactions with the stationary phase are minimized hereby, which is important for phases used in the ZIC-HILIC system, for instance. Secondly, hydrophilicity of charged residues can be decreased by minimizing hydrogen bonding when protonating them. Except for sialic acid, oligosaccharides on the peptides are not affected by TFA and therefore its use is an effective way for discrimination from nonglycosylated peptides. In addition, for TSKgel Amide-80 stationary phases it has been found, that the usage of ion pairing reagents like TFA helps to minimize interaction of positively charged residues with silanol groups on the silica surface [48].

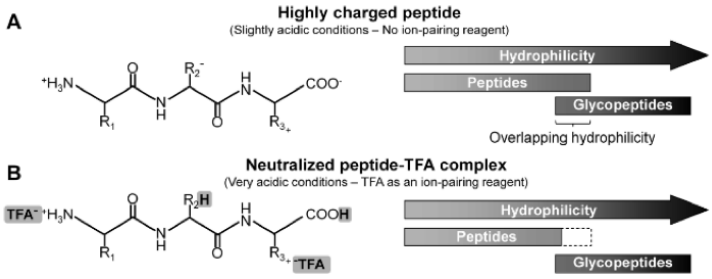


Figure 10: Ion-pairing effect of TFA in HILIC [47]

1.5 Mass spectrometry

Becoming the important tools in biological sciences that it is today, mass spectrometry (MS) has a long history of development, starting with J.J. Thomson's discovery of the existence of mass-to-charge (m/z) ratios in 1898. Even in modern mass spectrometry, Thomson's finding as well as the principle on which MS is basedon still hold true. In a first step, the molecules must be ionized in a gas-phase transition. Based on their m/z values, the ions can be separated using a mass analyzer, that generates either an electric or a magnetic field. In the last step, these ions can be detected at a specific time and their mass and abundance is recognized [49]. To employ this method on today's demands, where analysis of complex biological samples, protein structures, large protein networks, as well as medical applications and high throughput technology is required, there has been a constant drive in development of these three MS key steps.

Mass spectrometry-based proteomics has evolved as the key technology for large scale proteome research, as well as in targeted and quantitative approaches. Recently, mass spectrometry has also been used for the generation of structural proteomics data and large protein interaction networks [50]. In order to keep technical errors to the minimum and to maintain a high degree of reproducibility, a simple and fast workflow is demanded, which especially for analysis of post translational modifications (PTMs) remains challenging. In typical proteomics experiments proteins must be purified in their according tissue or cell suspension. Different techniques for cell lysis are commonly used, including detergent-based lysis, chloroform/methanol extraction or high-pressure lysis [51]. In bottom up approaches enzymatic digestion prior to MS analysis is necessary to generate a more homogenous analytical species, using endoproteinase like trypsin, Lys-C or Glu-C, [52]. Enrichment strategies can be applied, accordingly to the PTM of interest, combining a large range of methods, such as chromatographic separation, SDS-PAGE and affinity enrichment. In MS analysis, the identified peptides are then compared to a protein database. This process has been further optimized, in the introduction of shotgun proteomics, where even smaller peptide ions are generated. Those peptide fragments can be mapped to precursor ions using MS/MS search engines like SEQUEST, MASCOT or MS Amanda. To consider false positive matches, decoy databases are used that help to define a probability cutoff, that separates false positive from actual positive peptide matches [37].

1.5.1 Ionization

In the context of biomolecule analysis electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) have emerged as main ionization methods. Both are examples of soft-ionization, meaning that primary ions with low internal energy are generated. Further fragmentation is hereby reduced, in contrast to hard ionization, such as electron ionization (EI) [53]. In addition, there is fast atom bombardment (FAB), secondary ion mass spectrometry (SIMS) and atmospheric pressure chemical ionization (APCI).

MALDI ionization principles are based on direct formation of ions out of a solid phase by photons, mostly using a pulsed 337 nm nitrogen laser. During sample preparation, the analyte is embedded in a solid matrix of organic molecules that absorb at the wavelength of light emitted by the laser. Co-crystallization with matrices, such as 2,5-di-hydroxy benzoic acid or derivates of cinnamic acid, is mainly done on MALDI sample plates, that are exposed to a vacuum system [54]. However, the samples can also be ionized at atmospheric pressure and accelerated towards the mass analyzer, like in time-of-flight (TOF) instruments in MALDI approaches [55, 56]. Formation of single protonated analyte ions during protonation process is essential for MALDI, although the exact mechanism is still under debate.

Two models have been established to describe the protonation mechanism, the "Lucky Survivor Model" and the "Gas Phase Protonation Model" [57]. The original lucky survivor model proposes preservation of native charges of the analyte during co-crystallization with a matrix [58]. Normal

conditions in protein biochemical workflows include acidic sample preparation, assuring a positive charge on the peptide or protein, in combination with the respective counterion. The first possible outcome of this model is dominated by an overall neutral charge, achieved by laser ablation of ion clusters with the same amount of their according counter-ion. The second measurable possibility is the formation of charged ion clusters, that are generated during charge separation by laser induced disruption of the analyte-matrix crystal. In this case, counter-ions are removed and a positive charge is dominant. They can be accelerated towards mass analyzer and be detected as "Lucky Survivors". In the gas phase protonation model, neutral analytes are initially formed either by originally uncharged species in the matrix or by counter-ion attachment [59]. Ionization is achieved by collision induced proton transfer in gas phase to analytes from either protonated or deprotonated matrix ions.

In terms of sensitivity, MALDI has the advantage of a pulsed laser, offering a contemporaneous ionization event for analytes. Together with the uniformly charged ions and the usage of a mass analyzer like TOF, a high sensitivity can be achieved, thus little amounts of sample are needed. One disadvantage is the chemical noise, generated due to the matrix, causing limitations in the size of measurable analytes [56].

Another soft ionization method that is operated at atmospheric pressure, electrospray ionization, is based on the generation of a continues ion spray. Analyte in solution is guided through a small diameter capillary and accelerated by an applied potential difference towards a counter electrode. Often a concentric flow of gas, such as N_2, is used to support formation of this aerosol. With respect to polarity of the used voltage, either negatively or positively droplets are formed, that contain both highly charged analyte molecules and solvent [56]. Two forces, the surface tension and the electrostatic coulomb potential, dominate the electrospray process. Whereas surface tension tries to be minimized by transferring the liquid in an ideal state and hold it back in the capillary, Coulomb attraction forces it in the direction of the applied potential. This counteraction results in an elliptic arrangement of the liquid [60]. At the point, where the electrostatic attraction (Taylor Cone voltage) gets larger than the surface tension, liquid shapes into a cone (Taylor cone), from which's tip a droplet containing spray is generated [61].

Two ionization models have been supposed for electrospray ionization. The ion evaporation model postulates, that due to evaporation of the solvent, droplet size is decreasing during acceleration. Below the Rayleigh limit, but at a critical field strength at the droplet surface, where the energy can compensate for surface enlargement, solvated ions get expelled from the droplet when reaction free enthalpy is reached. In this model, the generation of large, randomly charged droplets is suggested, due to distinct evaporation rates of ions with different chemical and kinetical properties [62]. In contrast, the charge residue model is based on the idea, that only one ion species containing droplets are generated during electro spraying, where ionization is independent of the ions. Like in the ion evaporation model, the solvent evaporates over time and ions are released during this process. Evaporation efficiency is the driving force in this model, whereas ion properties play a minor role. The droplets emitted from the original Taylor Cone decrease in size during evaporation. As soon as

a critical field strength is reached, and the Coulomb repulsion overcomes the surface tension, the droplets starts to form Taylor cones themselves. In a repetitive process, this mechanism is going on until the droplets are limited to one analyte ion each [62].

1.5.2 Fragmentation

With increasing interest in the analysis of large biomolecules like proteins, the development of tandem mass spectrometry was a large improvement in the field of proteomics. In MS/MS approaches, precursor ions can be isolated using quadrupole or linear ion trap technology and further fragmented. Fragment assignment to the according precursor ion is the mechanism for its characterization. Post-source fragmentation methods, like collision-induced dissociation (CID), higher-energy collisional dissociation (HCD) or electron-transfer dissociation (ETD), in combination with soft ionization had proven to be reliable fragmentation methods for protein identification. They became of great interest again, when post translational modifications (PTM), like phosphorylation and glycosylation, came into spotlight in proteomics research.

Collision-induced dissociation

CID relies on acceleration of ions in gas phase towards an electric potential, to a higher kinetic state and the collision with neutral gas molecules like helium, nitrogen or argon. During this inelastic collision, kinetic energy gets partially converted to internal vibrational energy which leads to bond breakage and fragmentation into product ions. The fragmentations mechanism depends on the precursor as well as on the used energy. Single-collision or "fast" CID is operated at high kinetic energies in the kiloelectron volt range [63]. However, low kinetic energies between 1-100 electron volt generally induce multiple collisions and are referred to as slow CID method (10-100 µs between collisions). In slow-activation mode, in which CID is used in collision cells, such as in quadrupole or generally multipole region containing mass spectrometers, precursor ions undergo chemical changes and can dissociate or isomerize multiple times throughout the entire activation period [64]. Slow-heating methods that use "very slow" activation are characterized by the occurrence of competitive deactivation, such as cooling collisions or photon emission events. Ion trap based (Quadrupol ion trap i.e.) instruments are examples for slow-heating CID. Boltzmann distribution of the ions internal energy can be assumed, when dissociation rate is lower than the de/-activation rate and a "rapid energy exchange" is going on [65]. In peptide analysis, this distribution of internal energy using pathways of lowest energy results in the cleavage of the amide bond and leaves fragments of the b- and y-ion series [Fig. 11]. In the "mobile proton" model, fragmentation is explained by mobilization of a proton, localized at any basic site, by collisional activation towards backbone heteroatoms like carbonyl oxygen or amide nitrogen. Other examples for slow-heating CID methods are sustained off-resonance irradiation (SORI), Fourier transform-ion cyclotron resonance (FT-ICR) and blackbody infrared radiative dissociation (BIRD).

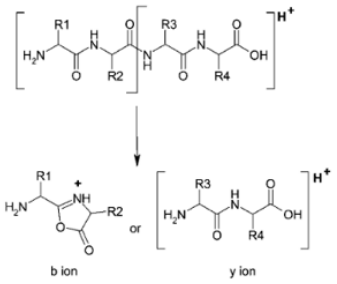


Figure 11: b- and y-ion series generated upon cleavage [65].

Higher-energy collisional dissociation (Higher-energy C-trap dissociation)

Another slow activation technique differing from conventional CID in its location as well as in its fragmentation is the "Higher-energy collisional dissociation" or originally "Higher-energy C-trap dissociation" (HCD). It is operated as beam-type CID, where a beam of ions is directed onto a collision region for excitation only during the passaging time in the collision cell. High dissociation rates are achieved by this, whereas ion deactivation rate is reduced, in contrast to ion trap-based CID [66]. HCD first was applied on an LTQ Orbitrap, in which radiofrequency was increased to 2.500 V, to retain ions in the C-trap, that acted like a collision cell [Fig. 12].

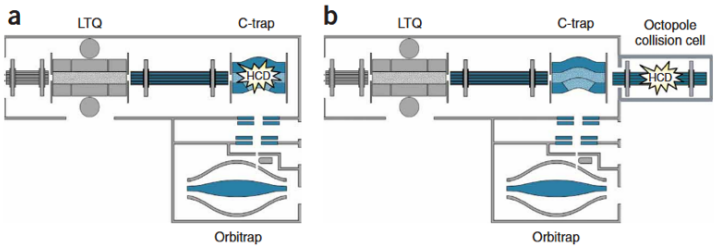


Figure 12: Comparison of traditional C-trap operated HCD cell and extension of LTQ by a external collision cell [67].

Due to the C-traps limitations in the capturing of low mass regions, when operated at high radiofrequencies, the instrument was extended by an octopole collision cell, which is located after the C-trap. The cell being controlled over radiofrequency voltages to change its current, is attached to a gas supply and operated under pressure with target gas like nitrogen [67]. Slow activation using HCD produces higher internal energies and enables additional, high energy pathways for fragmentation leading also to a- and z- ion series [65]. Importantly, this instrumental arrangement allows for identification of ions in the low-mass range of a mass spectrum, where a2 and b2 ions pairs as well as the y1 and y2 ions are located [67, 68].

Electron transfer dissociation

Electron transfer dissociation (ETD) originated from the electron capture dissociation (ECD) technology, developed for Fourier transform ion cyclotron resonance mass spectrometry. ECD is based on the peptide backbone fragmentation by thermal electrons, which produce hydrogen radicals, that add to carbonyl groups and cause fragmentation in a c- and z-ion type matter [Fig. 13]. Post translation modifications are thought to be preserved in particular by this sequence independent fragmentation

mechanism [69,70]. Restriction to FTICR instruments for ECD due to the instability of thermal electrons in radiofrequency operated instruments led to the development of a technology with broader application combining ECD and chemical ionization (CI).

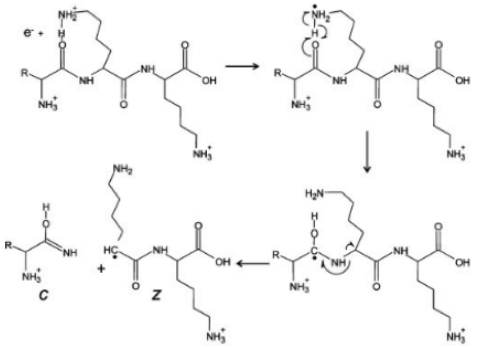


Figure 13: Fragmentation mechansim for production of c- and z-type ion series [69].

Electron-donors with low electron affinity, like anthracene and later fluoranthene, are used for exothermic electron transfer onto protonated peptides. In a similar fragmentation mechanism like ECD, producing a hydrogen-radical via an ion/ion reaction, the fragmentation pattern is also comparable with the production of c- and z-ions [71]. Low fragmentation efficiency in comparison to ECD, had been overcome by the usage of complementary methods, such as supplemental activation ETD (additional collisional activation) or activated ion ETD (activation by infrared laser) [72,73].

1.5.3 Mass analyzer

For selection and identification of certain masses or mass ranges, different approaches have been developed, all relying on the separation of ions, based on their mass-to-charge ratio. In combination with different ionization and fragmentation technologies, selection can be achieved at different levels during MS analysis. In a first step, precursor ions, that are ionized in the ion source, can be selected in an MS1 scan, either for a targeted approach or in a shotgun mode. In proteomics, different targeted methods are used to selectively detect specific peptides and further analyze the proteins. Selected reaction monitoring (SRM), multiple reaction monitoring (MRM) or parallel reaction monitoring (PRM) can be used to determine target peptides (precursor ions). At MS1 level, they get isolated and their "transitions" get analyzed (fragment ion-precursor ion pair) at MS2 level after fragmentation [74]. Shotgun and bottom-up proteomics aim for the identification of the whole proteome by selecting all tryptic peptide ions in MS1 and matching them to their fragment ions in MS2. Shotgun strategies have been mainly operated in a data-dependent acquisitions (DDA) mode. In DDA, precursor ions are selected with decreasing intensity and sent for fragmentation [37].

In the last few years, data-independent acquisition (DIA) strategies have been developed, where in contrast to DDA, sequential acquisition cycles are used like in SWATH-MS or BoxCar [75, 76]. Low abundant precursor ions can be detected as well without losing information, since DIA can mea-

sure all transitions between precursor ions to MS2 fragment ions. Over a defined mass range, the instrument cycles through specific m/z-wide windows (20 m/z e.g.) and comprehensively acquires both MS1 and MS2 spectra. Individual peptide elution peaks can be generated, when following the according transitions over the LC-MS/MS run, which allows for a targeted application as well. However, generation of large, organism or cell type specific databases is necessary [77].

Several mass analyzers can be used for these proteomics approaches in many different applications, either as standalone version, in tandem or in different combinations at MS1 and MS2 level. As one of the first of its kind, a time-of-flight (TOF) mass analyzer was operated in combination with an incontinuous method such as the MALDI ionization technology. Separation principle is based on the different flight times in a drift region, after ion acceleration to equal kinetic energies toward a potential. Flight times can be described by the square root of the ions masses [78]. Pulsed ionization methods had resulted in limitations for resolution since pulsed laser ionization leads to different initial kinetic energies. This time lag, due to ions, that get accelerated in different directions, at slightly different time points and at different energies was overcome by the introduction of time lag focusing or delayed extraction. Focusing is achieved by the usage of an acceleration voltage, equalizing differences in kinetic energies and velocities for ions of the same mass [79]. Another progress in the increase of resolution was the elongation of flight time by using reflectons, beeing able to reflect incoming ions and send them back to the detector, which has been mainly operated as photomultiplier [80]. However, one of the biggest advantages of TOF is its capability to analyze large molecules, with almost no restriction in m/z for biomolecules. TOF instruments have been used and adapted for different combinations, with ESI or MALDI as ion source, in tandem with quadrupoles or as TOF-TOF [81].

The idea of multipoles as mass analyzer or traps, has been realized in several technologies. They are using either a three-dimensional ion trap or a two-dimensional, linear ion trap. In a 3D ion trap, ions get captured in a system of two hyperbolic cap electrodes being faced at each other. The cap electrodes are combined with a hyperbolically ring-shaped electrode, located between those two electrodes. Ions that enter the trap are kept in an electric field, which directs ions towards the center, by a switching, oscillating field (radio frequency) and a static field [82]. However, 3D ion traps have limited storing capacities and therefore low mass accuracy. In a continuous flow, using ESI f.i., much information on peptides, that are ionized during the scanning event gets lost. To keep up with this continuous flow, linear ion traps or "2D ion traps" have been developed in linear trap quadrupole (LTQ) instruments, that rely on quadrupole as electric field geometry [37]. Quadrupoles can be used as ion trap, but also application as mass filter is possible, which is realized in many hybrid MS approaches using qTOF instruments or triple quadrupole systems. Similar to 3D ion traps, quadrupoles consist of hyperbolic rods, with two additional axial rods instead of a ring electrode. It is operated by creating a radio frequency field in 2D, that switches between each the two electrodes. In addition, it is supported by an axial DC voltage, which is supposed to keep the ions in the center of the quadrupole. Trapping efficiency could be increased with the use of linear traps, since ion can be easily taken up in the trap without a RF field in axis direction. Moreover, linear geometry

provides a much larger volume and therefore a higher ion capacity [83]. Triple quadrupole systems are widely used for continuous flows and especially targeted approaches like SRM and MRM have been adapted for such instruments. The instrumental set up includes three quadrupoles in series, that are named as Q1, Q2 and Q3. Q1 and Q3 are operated as mass filter using RF as well as DC, whereas Q2 is used as collision cell for precursor ion fragmentation and only uses RF [84]. Several scan modes have been established for triple quads, which include precursor scan, product ion scan and neutral loss scan. In these modes, the operator can decide whether Q1 or Q3 is operated in scan mode, or if a fixed ion value is set, or even scanning is performed in both quadrupoles, to detect neutral loss ions. Selected or multiple reaction monitoring methods make use of this possibility to filter fixed precursor ions in Q1 and then scan for each transition in Q3, either for one selected ion or for multiple ions [85–88].

The orbitrap was developed based on the idea of Fourier transform ion cyclotron (FT ICR) using Fourier transformation to translate oscillations into mass spectra. In the orbitrap, orbital trapping of ions in an electrostatic field around a central spindle electrode, that is coaxial with an outer electrode, is used for separation. The two electrodes are known to consist of a quadro-logarithmic electrostatic potential, in which the ions are forced to rotate around the central electrode and move along its z-axis with a frequency dependent of their m/z ratios. This harmonic oscillation is independent of the ions energy and their spatial spread. Mass spectra are generated measuring these frequencies in the form of current transients and converting them into a frequency mass spectrum [89]. In the first orbitrap containing instrument, the LTQ, ions are directed into the orbitrap by a device called C-trap. C-trap consists of C-shaped hyperbolic rods and two lenses with ion apertures, to close the system. Nitrogen gas is generally used for collisional damping. Cooled ions are trapped along the curved axis between the C-shaped electrodes, by applying a voltage to gate and trap electrodes. Using ramped voltages to create potential differences the ions are accelerated at high kinetic energies towards the orbitrap [89]. Several machines have been developed using orbitrap systems, such as the LTQ-orbitrap, the Orbitrap Elite, the Q Exactive and the Lumos Fusion. Advantages of the orbitrap technology are its high mass resolution (up to 150.000) with a large ion capacity, the high mass accuracy and high mass- and dynamic range [90]. Moreover, targeted approaches like parallel reaction monitoring (PRM) have been adapted for orbitraps in combination with computational approaches like Skyline, which reduces individual product ions scans in SRM/MRM to a single MS2 scan at higher resolution [91, 92].

1.5.4 Quantification

Mass spectrometry itself is not a quantitative tool, which is why several different approaches have been established, helping to add the quantitative component to comparative analysis in protein research. Relative quantification can be achieved by either using mass tags in the form of differential isotopic labelling techniques or by label-free approaches. Absolut quantification can also be done, using internal standards like synthetic peptides [93, 94].

Two different approaches are currently used, that do not rely on isotopic labelling, but on the correlation of the mass spectrometric signal and the protein quantity. One possibility is the direct comparison of extracted ion precursor spectra according (extracted ion chromatogram - XIC) to their signal intensity, determined either by chromatographic area or peak intensity. Another method uses counting and comparing of fragment spectra, identifying peptides and correlation to their relative protein abundance [95]. Label-free quantification is considered as the least accurate quantification method, since the amount of variation during the experimental procedure, either systemic or non-systemic, is just higher compared to labelling techniques. It is suggested, that label-free quantitative proteomics (LFQP) experiments are kept simple, with a low number of experimental steps, which could increase variance in the data. Strategies for PTM analysis, generally demanding for high effort in sample enrichment, are favorably performed with isotopic labelling. However, LFQP comes with the large advantages of being comparably cost-effective, with no labelling reagents necessary, and its fast and simple sample handling. Also, there are no theoretical limits in sample number, which can be compared with such approaches [96].

Metabolic labeling is the quantification method introducing least variance since labelling occurs already at the step of cell culturing. In stable isotope labeling by amino acids in cell culture (SILAC), heavy amino acids are metabolically incorporated into the cell during cell growth and cell division [97]. The moment of uniting different cell cultures for comparative analysis is the earliest possible, allowing unified sample handling throughout the whole proteomics workflow. SILAC uses 13C6- or 15N4-arginine and lysine, which are supplied by the medium, aiming for a peptide population that carries at least one labelled amino acid, after tryptic digestion. Quantification is achieved by relative comparison of signal intensities for a maximum three conditions (unlabeled, 13C6 and 13C615N4) of cells in culture, containing either "heavy" or "light" peptides. There have been attempts of applying the system onto higher organisms, but due to relatively high cost and time investment, SILAC is restricted to immortalized cell lines for higher eukaryotes [96].

One of the most frequently used technology for quantitative proteomics is the isotopic labelling of amino acids in peptides to reduce variance between runs, using different chemical approaches. Originally, isotope-coded affinity tag (ICAT) was developed as a covalent label, which could be used to specifically target cysteine residues. ICAT consists of a thiol-specific reactive group (iodoacetamide e.g.), a linker, which due to deuterium incorporation enables mass shifts to 8Da and a terminal biotin tag, for peptide avidin enrichment [Fig. 14] [98]. Cysteine selectivity is thought to help de-complexing samples by pulling only a homogeneous pool of cysteine containing peptides. However, ICAT is not feasible for quantification of large numbers of peptides on the other hand, since cysteine missing proteins are not detected at all.

As an alternative to target a broader range of peptides, isobaric reagents have been developed, that specifically bind to peptide N-termini and the epsilon group of lysines. In "isotope tag for relative and absolute quantification" (iTRAQ), as well as in "tandem mass tags" (TMT), N-hydroxysuccinimide (NHS) chemistry is used for covalent labelling [Fig 15] [99,100]. A big advantage of isobaric reagents

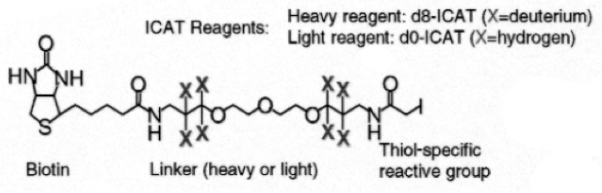
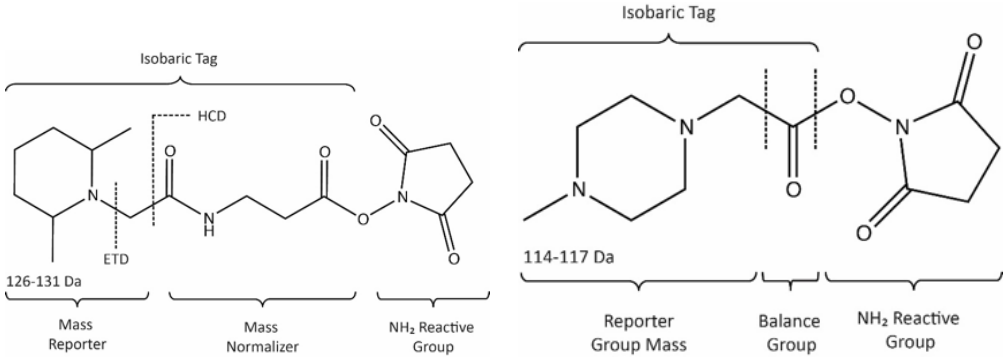


Figure 14: ICAT reagent for chemical labelling [98]

is elution profile, which does not change for different mass tags with identical chemical behavior. At MS1 level, individual multiplexed peptides appear as single peak with the same m/z. TMT contains both an ETD and HCD (originally only CID) cleavage site, between the mass reporter and the mass normalizer. During MS2, the mass reporter, get released and can be detected in the MS2 spectra, where a difference of each 1Da can be observed.



(a) TMT (b) iTRAQ

Figure 15: TMT and iTRAQ reagents for chemical labelling

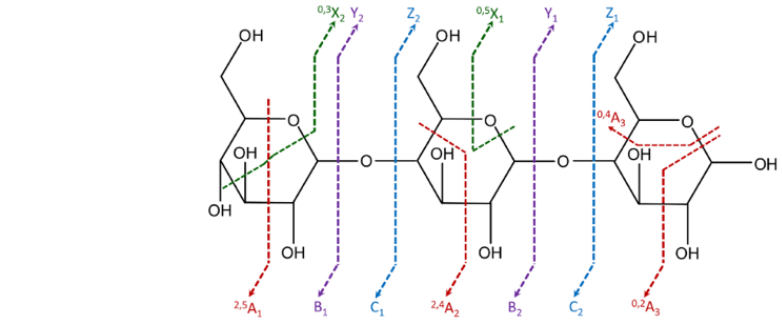
1.6 Glycoproteomics

Protein glycosylation has evolved as one of the most frequently analyzed post translational modifications due to its involvement in a broad range of biological processes and diseases and its increasing role as analytical biomarker. Consequentially, the field of glycoproteomics is growing with the aim of developing comprehensive strategies to analyze whole glycoproteomes. This kind of analyis can give information on the exact type of glycosylation as well as the precise position of glycosylated residues on proteins. Keeping into account the large dynamic range glycoproteomics is facing with, method development is driven towards pre-mass spectrometric sample enrichment, as well as fragmentation techniques and most importantly the bioinformatic data analysis and interpretation.

To gain information about biological relevance of glycoproteins and to determine their role in different parts of an organism and in distinct biological pathways, precise mapping of glycosylation sites is mandatory in large proteome studies. Also, the characterization of intact glycan structures, that can explain cellular localization of proteins, involvement in signaling pathways, as well as general cell response and cell homeostasis, is helpful. To maintain structural information, ionization, as well as

fragmentation mechanisms must be considered, that affect both glycans and the peptide backbone. MALDI and ESI can be used as ionization method for glycoproteomics approaches, although sialylated residues have been found to be unstable and prone to decay during MALDI ionization. ESI in contrast, allows for analysis of sialic acid containing glycopeptides and therefore the characterization of neutral and charged glycan residues [101].

Fragmentation techniques that can be applied for glycoproteomics reflect technology, which is normally used in proteomics MS/MS approaches as well. CID and HCD are the fragmentation methods that are mostly applied in N-glycopeptide analysis. CID, which generally induces amide bond cleavage due to the lowest energy barrier of dissociation, preferably cleaves glycosidic bonds in a carbohydrate B- and Y-type cleavage [Fig. 16a]. Additionally, oxonium ions of different glycan species appear in the lower mass range like Hex^+ (163.x Da), $HexNAc^+$ (204.1 Da), $NeuAc^+$ (292.1 Da) and $LacNAc^+$ (366.1 Da) oxonium ions [Fig. 16b]. Although generation of such signature ions, generated during ionization (ESI e.g.), can be helpful to produce elution profiles of tryptic glycopeptides, in-source fragmentation is generally preferred to be reduced when it comes to relative quantification [101]. Especially low abundant glycopeptides profit from the monitoring of fragmentation originating oxonium ions, which enable a high level of sensitivity.



(a)

(b)

Figure 16: a) Carbohydrate cleavage specificities [101]. b) Oxonium ion formation of glucose residues

Further development of mass analyzer technology and the introduction of the HCD fragmentation in combination with the orbitrap, impacted on glycoproteomic analysis as well. CID, which has mainly been operated with an iontrap mass analyzer, suffers from a cutoff in the lower mass range, where most of the diagnostic oxonium ions are present. Using a C-trap with a high resolution orbitrap helped to overcome this limitation. Beside formation of CID comparable oxonium ions, HCD leads to a pattern of smaller fragments of mainly mono- or disaccharides [102]. The most abundant ion

species in both CID and HCD spectra are the so-called Y1-ions [Fig. 16], which comprise peptides with a single GlcNAc. Detection of Y1 ions can be used to identify the glycan carrying peptides, which in combination with the glycan fragmentation pattern helps to identify the glycan structure [Essentials of, 3rd]. Due to the preferred fragmentation of the carbohydrate residues over the peptide backbone, insufficient information on peptide composition is generated with methods relying on the vibrational excitation and cleavage by collision with a neutral gas [103].

An alternative approach for peptide fragmentation is provided with electron-transfer dissociation (ETD) technology, enabling a higher peptide sequence coverage by generation of c- and z-type peptide ion series. No vibrational excitation is taking place during this process, hence post translational modifications such as glycosylation are not directly affected and resistant to cleavage during the fragmentation event [102]. However, ETD and other electron-based methods like ECD clearly have their limitations, especially when it comes to precise structure determination of glycans. Moreover, higher charged peptides are mandatory for such approaches, with a high charge state and low m/z values. Double and triple charged peptides, that are quite frequent in tryptic digests, that appear above 1.400 m/z are not well affected by ETD, but only lead to non-dissociative electron transfer [3, 102, 104]. Approaches to increase the charge states on the peptide population include chemical derivatization, like TMT-labelling, as well as the usage of alternative digestion strategies with LysC and GluC [3].

Extension of the typical tandem MS workflow by one dimension, performing a MS3 scan, has given good results on glycopeptide samples [105]. Either with combinations of different CID/HCD energies at MS2 and MS3 or by mixing CID/HCD (MS2) with ETD (MS3), it is possible to generate both fragments on glycan and peptide level and to target charged-reduced precursors [106, 107]. Using stepped normalized collision energy (NCE) to fragment both glycan and peptide backbones in one spectra, is a widely applied method for intact glycopeptide identification [108, 109]. Oxonium ion detection at MS2 level during HCD fragmentation had been used as trigger for ETD fragmentation in a targeted way to reduce the number of non-glyco spectra, generated in the dual approach [110]. In a newer approach, ETD and HCD are combined in a hybrid dissociation method called EThcD [111]. In contrast to methods using MS3, unreacted and charge-reduced precursor peptide ions are considered for further fragmentation and not removed after MS2. In one spectrum both ETD and HCD originating ion species (c/z and b/y ions) as well as glycan fragments are visible [105, 112].

Several bioinformatic tools, like BIONIC, GlycoFinder and SugarQb have been developed over the last years, to enable data simplification and selective glycopeptide identification [113, 114]. SugarQb is a glyco-node, which has been implemented into the Proteome Discoverer software [115]. It comes with a series of tools, enabling charge deconvolution and de-isotoping, prior to glycan identification, to generate spectra that contain only singly charged ions. Intact peptide ions, with glycans attached at different fragmentation state, are shifted to higher m/z ranges by this process, which makes less abundant peptide fragment ions in the low mass range more prone for detection. In between these regions, the dominant Y1 ion (peptide + $HexNAc^+$) appears [Fig. 17]. Identification of this singly charged Y1 ion is achieved by sequentially subtracting all potential glycan masses considered ac-

cording to an established glycan database, from the original precursor ion and assuming it as new, modified precursor ion. Peptide identification can then be performed on the cloned spectrum with a new precursor, using any established MS/MS search engines, with "HexNAc" as variable modification on asparagine, serine or threonine.

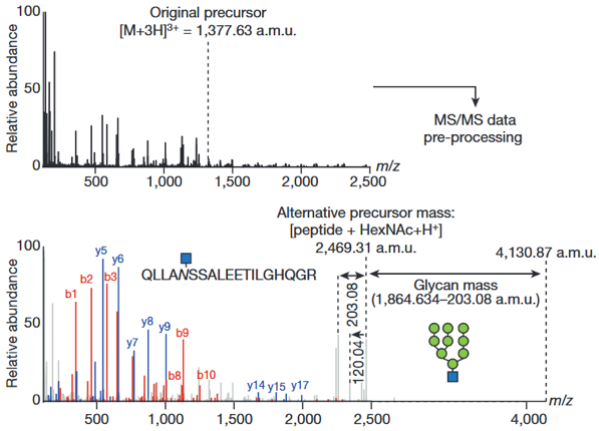


Figure 17: SugarQb precursor detection and determination of an alternative precursor mass [Stadlmann et.al., 2017]

1.7 Aim of study

Catalytical subgroup isoforms STT3A and STT3B of the OST complex have been largely studied and described in terms of molecular function and target specificity. Separation into post-translational and co-translation glycosylation and their compensatory role of each other has been accepted by the community. This knowledge has mainly been generated investigating glycoprotein substrates in mammalian cell model systems and bioinformatics analysis of proteomics data sets. STT3A is the key enzyme in protein glycosylation, whereas STT3B plays a supportive role to ensure folding of proteins with unfavorable glycosylation sites. Knockout of the main player driving protein glycosylation has been found to trigger unfolded protein response and is expected to challenge the less active and selective co-translational isoform. However, cell that lack STT3A are still viable and the question, how the compensatory effect of each isoform is reflected at glycoproteome level, still remains. In this study, a quantitative glycoproteomics wokflow is developed, aiming at the analysis of intact glycopeptides. A comparative approach is used to evaluate glycosylation levels as well as to get information on the intact glycoforms attached to peptides. This is achieved by using different chromatographic techniques for selective glycopeptides enrichment and better separation of the whole proteome. Bioinformatics analysis is performed using the in house developed glycosylation identification tool SugarQb, which is implemented into the proteome discoverer interface. Mouse embryonic stem cells with known expression data are used as model system to compare knockout of either STT3A or STT3B to a wildtype. Glycoproteomics data is used to investigate on the established specificities that are addressed to the STT3B isoform. Peptides containing closely spaced acceptor sites, N- or C-terminal domains or unfavorable middle-x amino acid residues are extracted and compared between the different states. Moreover, data is analyzed on peptides containing new motifs, that

may be enriched upon selective knockout, not only for STT3B but also for STT3A. Finally, biological consequences and phenotypic appearance upon interference with the glycosylation machinery are determined.

2 Material and methods

2.1 Chemicals

Acetonitrile, $a \geq$ b 99.9%, HPLC grade
Ammonium Acetate, $a \geq$ b 98%
Chloroform, $a \geq$ b 99,9%, HPLC grade
Disodium phosphate, $a \geq$ b 99.5%
1.4-Dithiothreitol
Ethanol, absolute
Formic Acid, 98%, MS grade
HEPES, cell culture grade
Hydrochloric Acid, 37%
Hydroxylamine, 50% wt. In H2O
2-Iodoacetmide, >99%, crystalline
Methanol, $a \geq$ b 99.7%
Monopotassium phosphate, $a \geq$ b 99.5%
PBS, cell culture grade
PNGase F from Elizabethkingia meningoseptica, $a \geq$ b 95%, for proteomics
Potassium Chloride, $a \geq$ b 99.5%
Sodium Chloride, $a \geq$ b 99.5%
Sodium Hydroxide, $a \geq$ b 99.5%
TMTsixplex^{TM}Isobaric Label Reagent
Triethylamine, $a \geq$ b 99.5%
Trifluoroacetic acid, $a \geq$ b 99%
Tris(hydroxymethyl)-aminomethan - HCL, $a \geq$ b 99.9%
Trypsin Gold, Sequencing Grade
Ultrapure Water, Milli-Q
Urea, reagent grade, 98%

2.2 Sample preparation

Cell lysis and protein digestion

Mouse embryonic stem cells were harvested and washed three times with PBS to get rid of remaining HEPES buffer and BSA, with centrifugation steps at 500 rcf, at room temperature in between. Cell lysis was performed in a mixture of chloroform and methanol (2:1) to achieve especially membrane glycoprotein extraction of the lipid bilayer. About 50ţl cell pellet was vortexed sorely in lysis solution and put in an ultrasonic bath for 10 minutes. After complete dissolution of the cell pellet, 100ţl ethanol was added and the pellet was dried almost to completion in a speedvac aperture. The pellet was then taken up in 450ţ ethanol with 50ţl water, mixed and rotated again until evaporation of the ethanol. In the next step, 500ţl 8M urea containing 50mM HCl was added to the sample and sonicated again for 10 minutes. With Tris buffer, the pH was adjusted to pH=8 in a volume of 600ţl. Reduction and alkylation were performed with 10mM DTT at 30 degree Celsius for one hour and 20mM IAA at room temperature for 30 minutes. Prior to the overnight digest with 10ţg trypsin at 37 degree Celsius, the sample was diluted to 4M urea with water. Lysis procedure and enzymatic digest was performed in parallel for each triplicates of the wildtype (two times) and knockout of either STT3A or STT3B.

Desalting and quality control

Digested samples were then desalted offline on Sep-Pak®C18 3 cc Vac cartridges (200 mg, reversed-phase, 55-105 µm particle size, 125Åpore size), to get rid of excessive salts, DNA and RNA remnants as well as undigested material. Before application onto the columns, samples were acidified with FA to pH2-3. C18 columns were first equilibrated with 2x 3ml methanol and 1x a mixture of ACN and 0.1% FA in water (1:1). After washing the columns with 3x 1ml 0.1% FA, the samples were applied and again washed with 3x 1ml 0.1% FA. Elution was achieved with 2x 600µl 50%ACN in 0.1% FA. In the speedvac, the samples were dried to remove the organic phase. A monolithic system was used to control the samples purity and concentration using a Dionex UltiMate 3000 HPLC RSLC nanosystem with a PepSwift Monolithicő RSLC column (0.2 x 5 mm, Thermo Fisher Scientific) at 60 degree Celsius. The system used a 20 minutes gradient from 2-90% buffer B (80% ACN, 19.92% H2O, 0.08% TFA), a 5 minutes washing step at 90% buffer B and a 2 minutes gradient from 90-2% buffer B. Column equilibration was done at 2% buffer B for 13 minutes. Relative concentration determination using peak areas of the peptide elution profile was used to adjust the peptide concentrations of triplicates of both wildtype and knockout, prior to isobaric labelling.

TMT labelling

For the labelling procedure, each 150µl of each sample was diluted with 50µl HEPES pH8.5. Each TMT labelling reagent (126-131) was dissolved in 50µl ACN and then distributed on the samples (wildtype: 126-128; knockout: 129-131). The labelling step was done at room temperature for 1 hour. Reaction was stopped with hydroxylamine hydrochloride (10µl in 80µl H2O+10µl conc. HCl) and the pH was adjusted with TFA (conc.) to pH=3. 300µl of each sample were then pooled together and completely freed from organic phase on the speedvac. Again, desalting of the pooled

sample was done on a 200mg C18 SEP-PAK cartridge [Desalting and quality control] and eluted in 2x 600µl 50% ACN in 0.1% TFA in water. After desalting the sample volume was reduced to 100µl on the speedvac for further processing.

High pH chromatography

To generate a chromatographic orthogonality towards RPLC prior to mass spectrometric analysis and to decomplex the sample for proteomics measurement, a preparative chromatographic sample separation was performed at higher pH than generally used in a reversed phase system. On a Dionex UltiMate 3000 HPLC RSLC nanosystem with a Kinetex®50µm EVO C18, 100 Åcolumn (250 x 4,6 mm, Phenomenex), 10µl labelled sample in 90µl H2O at pH= 8.5 (adjusted with TEA) was loaded for separation. In a 30 min gradient at 1ml/min, the organic phase concentration was increased with buffer B (80% ACN, 19.9% H2O, 0.1% TFA) from 5-50% B. In 1 minute, buffer B is increased to 80%, at which it was kept for 5 minutes. Finally, buffer B was decreased to 5% again and the column was equilibrated for 10 minutes. The sample was fractionated into 30 fractions of each 1 minute collection time using an auto collector. Fractions were then pooled in a way, that iteratively the 1st to the 10th of each 10 fractions are pooled together, resulting in 10 final fractions for both STT3A and STT3B experiments. In the speedvac, the sample volume was reduced to 100µl.

HILIC

For selective glycopeptide enrichment, two types of preparative HILIC were performed, both using a Dionex UltiMate 3000 HPLC RSLC nanosystem with a TSKgel Amide-80®5µm column (250 x 4,6 mm, TOSOH BIOSCIENCE). In a first step, 40% of the sample was used to perform a batch enrichment of glycopeptides in the complex sample by loading in total 100µl of 10µl sample (reduced volume) in 80% ACN and 1% TFA to the HILIC column. In the first 20 minutes of the gradient, hydrophobic peptides where eluted from the column at a constant flow of 1ml/min of 75% buffer B (100% ACN, 0.1% TFA). Rapid decrease of ACN concentration to 50% forced hydrophilic compounds, mainly glycopeptide, to be eluted from the column. Fractions were taken for 5 minutes at 50% ACN, until the concentration was increased to 80% to wash the column. Batch enriched glycopeptide fractions were pooled and the volume was reduced to 20µl. Half of the sample (10µl) was diluted with 80% ACN and 1% TFA in water and applied to the HILIC column. A longer HILIC gradient was then used to separate the already batch enriched glycopeptides. In a 25 minutes gradient the ACN content was reduced from 80% to 50%, were it stayed for 5 minutes, before the column was equilibrated for 10 minutes. According the elution profile, 10 fractions around the latest eluting peptides were taken and the volume was reduced to 100µl. For both HILIC enrichment strategies TFA was used as ion pairing reagent to achieve a better separation by increasing hydrophobicity in non-glycan containing hydrophilic peptides.

PNGase F digest

The second half of the batch enriched, pooled glycopeptide samples (STT3A and STT3B) was used for a de-glycosylation experiment. In 500µl volume with 20mM TRIS pH 8 the sample was digested with 1U PNGase F (Sigma Aldrich) over night at 37 degree Celsius. The samples were acidified

with TFA to pH=2-3 and desalted using a 200mg C18 SEP-PAK cartridge. Afterwards, the sample volume was reduced to 100µl on the speedvac.

2.3 LC-MS/MS-analysis

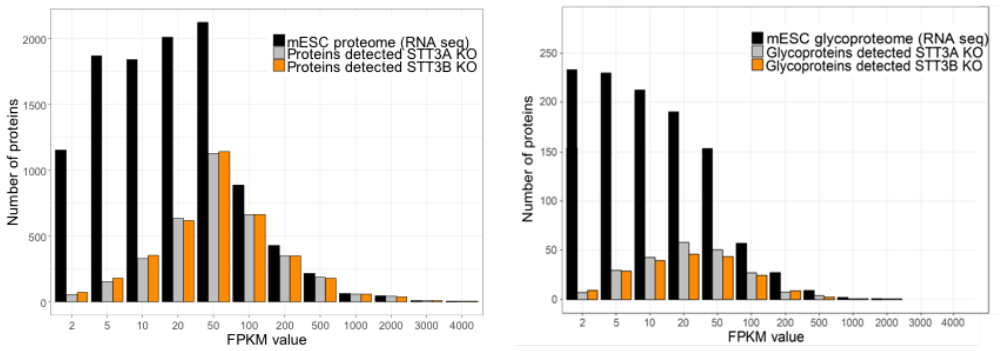
Prior to each mass spectrometric analysis, a Dionex UltiMate 3000 HPLC RSLC nanosystem was located upstream of the mass spectrometer. An Acclaim PepMap C-18 precolumn (0.3 x 5mm, Thermo Fisher Scientific) was used at 25/min with 100% buffer A (99.9% H2O, 0.1% TFA) for a first washing and desalting step, as protection for the analytical column. Reversed phase separation was operated on an Acclaim PepMap C-18 column (50 cm x 75 , Thermo Fisher Scientific) at 230 nl/min and 30 degree column temperature. Different methods were used for the glycoproteomics and proteomics workflow. HpH fractionated and HILIC enriched samples were separated by a method, starting with 10 min at 2% B (80% ACN, 19.92% H2O, 0.08% TFA), followed by a 60 min gradient of 2-35% B, a 35-95% increase in B, 5 min at 95% B, 2 min to decrease B from 90% to 2% and a final equilibration step of the column at 2% B for 17 minutes. For PNGase F digested samples, the gradient from 2-35% B was elongated to 180 min.

The Orbitrap Fusion^{TM}Lumos^{TM}Tribrid^{TM}mass spectrometer (Thermo Fisher Scientific) was used to perform all mass spectrometric analysis for TMT-labelled samples. In DDA mode, MS1 were generated in a 500-2000 m/z range, at an orbitrap resolution of 60.000 and a maximum injection time of 50 ms. Top N of 10 precursor ions were used to acquire MS2 spectra after HCD fragmentation at 38 $\pm$ 5% stepped normalized collision energy (NCE). Charge state 2-4 were allowed, as well as dynamic exclusion of precursors for 10 seconds after the first detection of a precursor at $\pm$ 10 ppm mass tolerance. The orbitrap was operated with 250 ms maximum injection time at 50.000 resolution, using a 1.2 m/z quadrupole precursor isolation windows and a fixed first mass of 115 Da. Knockout cells were determined by a PRM experiment on the Lumos and data evaluation was performed in Skyline.

3 Results

3.1 Mass spectrometric analysis

Using a high pH gradient on a reversed phase chromatographic system enabled equal distribution of peptides over the 10 fractions, which led to an improved elution profile and a high identification rate in the shotgun experiment. Each fraction covered between 2000 and 3000 unique peptides, which then could be further separated on a reversed phase column prior to MS analysis on a three-hour gradient [S 1, 2]. Using this approach, the identification and quantification of 4822 unique proteins (24579 peptides) in the STT3A KO experiment and 4851 unique proteins (22955 peptides) in the STT3B KO experiment was possible [Fig. 19a]. Glycopeptide-enrichment based on HILIC separation with an amide-80 chromatographic column combined with the bioinformatic peptide identification approach targeting glycosylated peptides with the SugarQb, led to the identification of 396 glycoproteins (1625 glycopeptides) in the STT3A KO experiment and 248 glycoproteins (1233 glycopeptides) in the STT3B KO experiment [Fig. 19a]. Direct comparison of the different experiments showed a large recovery of identified proteins between the high pH analysis, whereas recovery between HILIC runs of the different knockouts was quite low [Fig. 19b]. Comparison of proteomics data with RNA expression levels, using RNA-seq data, results in good coverage of highly abundant proteins. Less abundant proteins with RNA-seq FPKM values lower than 2 cannot be easily detected with this proteomics approach. Glycoproteomis analysis experienced similar limitation with bad coverage of low abundant proteins, that still can be targeted with methods like RNA-seq despite chromatographic enrichment strategies [Fig. 18].



(a) (b)

Figure 18: Comparison of mESC RNA-seq data with a) proteomics and b) glycoproteomics data based on their expression levels.

Comparative glycoproteomics approaches are mainly limited due to the necessity of finding equivalent peptides in both glyco-enriched as well as the hpH-fractionated samples, to get clear information whether a specific site is up or downregulated compared to the whole protein. Since glycoproteins are often low abundant, the recovery of glycoproteins in a normal proteomics sample is quite small. Glycoproteomics data generated in this study was therefore treated individually and was not strictly coupled to peptide equivalents in the proteomics data sets.

Gene ontology enrichment analysis of the cellular components for glycoproteomics data of STT3A KO cells show only a slight variation in ER- and lysosomal glycoproteins. Induction of STT3B knock-out leads to an upregulation of plasma membrane associated glycoproteins. Lysosomal proteins, in contrast are downregulated [Fig. 19c].

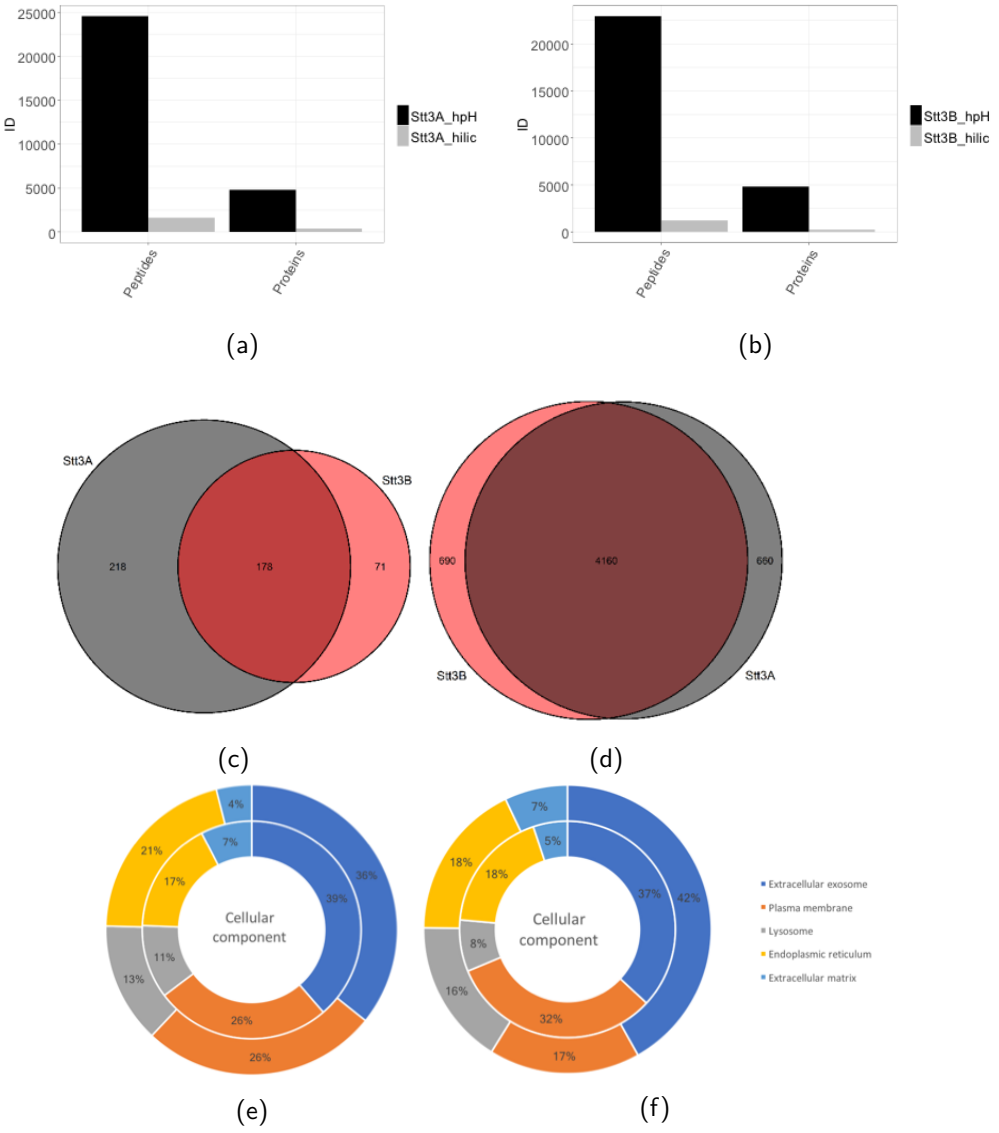


Figure 19: a) Identified proteins and peptides for both enrichment strategies in STT3A KO and STT3B KO cells. b) Overlap of glycoproteins vs. overlap of proteins between STT3A KO and STT3B KO. c) GO-analysis for both KO experiments - outer ring shows downregulated peptides; inner ring shows upregulated peptides.

3.2 Glycopeptide motif analysis

To answer the question if STT3B motif specificity can be detected on a HILIC enriched glycoproteomics data set of a STT3B depleted mESCs, different strategies were applied to analyze the samples on glycopeptide level. Intensity foldchanges (log2[intensity-foldchange]) of glycopeptides used for analysis were treated as mean values of peptides with different attached glycoforms.

Closely spaced acceptor site analysis

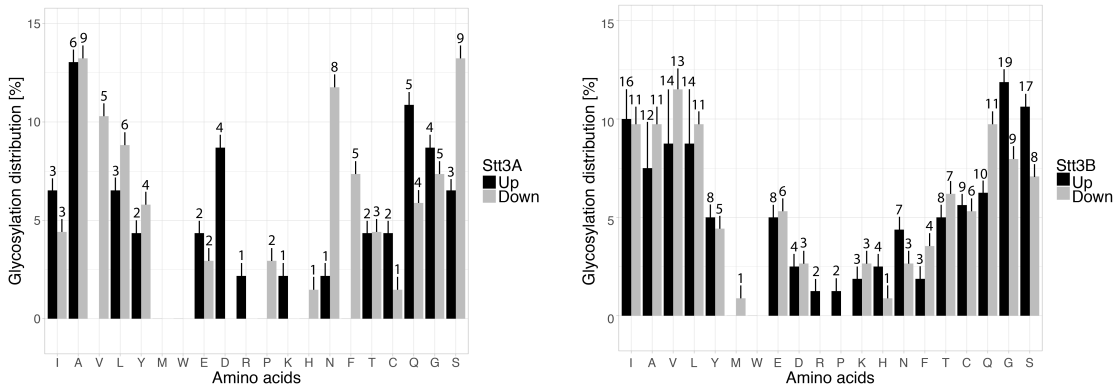
STT3B KO experiment data was manually checked for the occurrence and intensity change of closely spaced acceptor sites, to get insight on whether there is a STT3B dependence for such sites. In contrast, a potential dependence towards STT3A was analyzed. Out of 423 peptides carrying multiple glycan motifs, 4 peptides from 4 distinct proteins could be identified, that contained a motif with acceptor sites close to each other [Table 2]. They neither followed the proposed rule of NxTNxT or NxSANxS nor showed unfavorable amino acid composition, that could interfere with OST activity during glycosylation. Acceptor sites in Embigin (N216GS, N221ET), Lamp1 (N252DT), Tfrc (N725IT, N730ET) and Tor1aip2 (N318AS) were validated and the mean fold change of all glycosylation forms evaluated from both experiments. From all found motifs, no direct effect of STT3B on adjacent or closely spaced acceptor sites could be observed. Moreover, no influence on glycosylation efficiency of STT3A could be seen on closely spaced acceptor sites.

Gene name	Peptide	STT3B KO - fold change	STT3A KO - fold change
Emb	YIINGSHAnETR	+1.7	+2.4
Lamp1	AFNISPnDTSSGSCGINLVTLK	-1.2	-1.7
Tfrc	NITAFnETLFR	-1.9	-0.5
Tor1aip2	HLnASNPSEPATIIFTAAR	+0.4	-0.5

Table 2: Closely spaced and adjacent acceptor sites found in STT3A KO and STT3B KO datasets

Hydrophobic or negatively charged middle x-amino acids

Using an in-house developed motif finder software (QbToolkit), downregulated and upregulated glycosylation sites over the peptides were extracted from STT3B KO and STT3A KO data and analyzed for an enrichment of hydrophobic and negatively charged amino acids among the glycosylation motifs. From a pool of peptides from STT3a KO (upregulation: n=46, downregulation: n=69) and STT3B KO (upregulation: n=160, downregulation: n=113) experiments, glycosylation distribution was determined on motifs with different amino acids in the middle position (NxT/S/C) [Fig. 20]. According to the hypothesis, of STT3B acting as limiting factor for motifs containing hydrophobic and negatively charged amino acids, a downregulation of such motifs in the STT3B KO experiment in comparison to the STT3A KO could provide confirmation. Cut-off levels at $\pm$ 0.5 intensity foldchange and a probability of 0.05 for a significant change in glycosylation levels were used. Relative abundance of hydrophobic and negative amino acids appeared partially to be slightly less in cells lacking STT3B, although absolute quantities showed a different picture, when comparing a larger set of peptides showing upregulation, compared to downregulated peptides. On this glycoproteomics datasets no clear influence of STT3B was visible.



(a) (b)

Figure 20: Glycopeptides in STT3A KO- and STT3B KO cells that show a significant change in translation (black= upregulation, grey=downregulation). The number above the bars indicate the absolute quantities of identified peptides. Glycosylation distribution resembles the relative quantities.

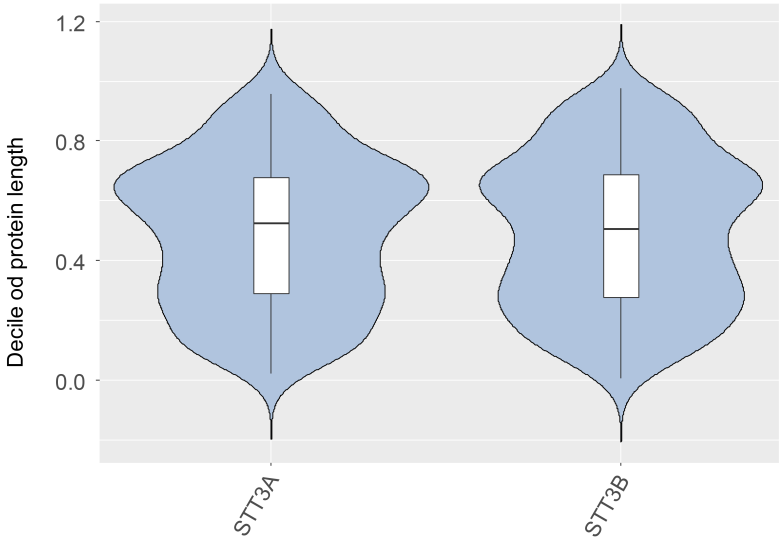
Glycosylation site distribution

QbToolkit was used for precise glycosylation site mapping by applying an algorithm that searches sequence specifically in the provided FASTA file output from the previous database search in Proteome Discoverer. The identified peptides were then mapped to the according protein sequence, the glycosylation motif NxT/S/C was centered, and the peptide sequence modified in a way, that it got prolonged or reduced to a sequence length of choice. Glyco-site distribution could be shown by calculating the decile based on the relation of glycosylation position and total protein length. Information that could be obtained from the distribution and localization comprised details on C-terminal glycosylation and glyco motifs close to the signal peptide sequence at the N-terminal domain of a protein. Evaluating a STT3B dependency for both C-terminal and N-terminal located glycosylation was limited given the glycosylation motif distribution, which was strongly centralized towards the middle decile of protein lengths [Fig. 21a]. Peptides concerned from terminal glycosylation were therefore quite rare and only a few could be observed in both datasets [Fig. 21b, c].

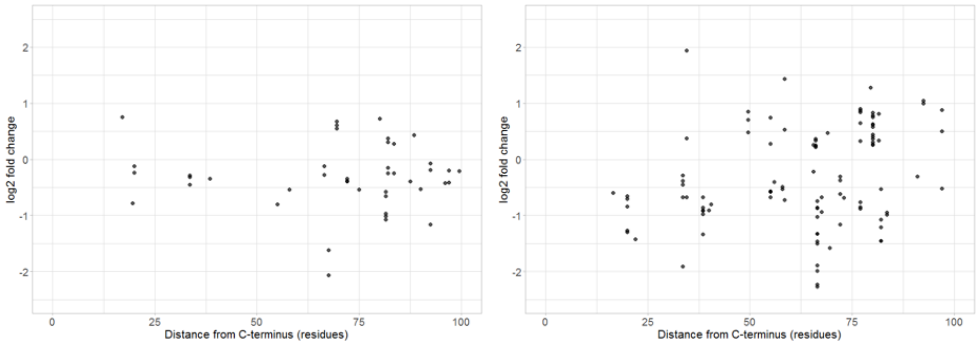
In the STT3B KO experiment, six proteins were found with a C-terminal glycosylation. A bias towards reduction in glycosylation level due to STT3B KO could not be found, with PSMs showing partially contrary glycosylation for different glycosylation forms of the same peptide. From these proteins, four could be recovered in the STT3A KO experiments dataset, showing similar results on peptide level. Moreover, the C-terminal distribution is a good approximation for the whole protein length [S 3, 4].

To evaluate a possible preference of STT3B for post translational glycosylation of acceptor sites close to the signal peptide, a cutoff of 50 residues from the N-terminal domain was chosen in the STT3B Ko dataset, since signal peptide length in eukaryotes can be found in a range of 15 to 40 residues [116]. Manually, the signal sequences from all concerned proteins were determined and its distances to the closest acceptor site were calculated. For any of the 10 proteins, for which peptides could be identified in this range, no acceptor site closer than 13 residues away from the signal peptide

could be found [Table 3]. Moreover, no decrease in glycosylation could be identified in this range for the STT3B KO.



(a)



(b) STT3A KO vs. WT (c) STT3B KO vs. WT

Figure 21: a) Distribution of glycosylation sites of the proteins identified in both knockout samples. b) C-terminal region of PSMs identified in STT3A KO cells. c) C-terminal region of PSMs identified in STT3B KO cells.

Gene name	Peptide	Δsignal peptide	STT3B KO -fold change	STT3A KO - fold change
Ctsc	SDInCSVMEATEEk	29	-0.6	NA
Emb	YSnLSLk	22	+0.7	-0.3
Hsp90b1	EEEAIQLDGLnASQIR	41	-1.1	+0.9
Il6st	GSnFTAICVLk	21	+0.4	-0.2
Mpzl1	EIFVVnGTQGk	15	-0.94	-0.2
Nup210	VnFTLEASEGCYR	18	+0.8	NA
Pgap3	CEERnCSGDALk	17	+1.0	-0.1
Ptgs2	TGFYGEnCTTPEFLTR	36	+1.2	NA
Pttg1ip	VGCSEYTnR	13	+0.7	-0.4
Timp1	FMGSPEInETTLYQR	30	-0.2	-0.05

Table 3: Identified secretory protein peptides containing glycosylation close to a signal peptide.

3.3 Translocon analysis

As part of the translocon complex, depletion of the catalytic subunit of the OST complex is thought to have an impact on the general composition of the translocon. Knockout approaches to selectively remove one of the distinct subunits have shown how translocon association influences the posttranslational and co-translational glycosylation behavior. In a mass spectrometric approach, this behavior should be reconstituted and the translocon composition of Stt3A and Stt3B depleted cells should be evaluated. Therefore, protein translation levels of members of the OST complex as well as SEC61, Tram1 and the Trap complex were compared between both knockout experiments.

Members of the STT3B containing OST complex (MagT1, Tusc3) were going down in their concentration, in case of detection, whereas members of the STT3A containing OST complex (DC2/OSTC, KCP2) could not be detected due to unfavorable protein composition for proteomics analysis and small size. Interestingly, members of the OST complex that occur in both isoforms were downregulated in the STT3A depleted isoform and upregulated in the STT3B depleted isoform. Furthermore, translocon associated proteins, such as the Trap-complex proteins Ssr1-4 were upregulated in the STT3B KO samples and vice versa in the STT3A KO. STT3B KO cells, however, seem to compensate their glycosylation levels by increasing production of Sec61, Trap and Tram1, as well as by increasing levels of STT3A containing OST complex to a large extend. STT3A KO cells, in contrast, were only slightly or not (Trap complex) affected by an increase in proteins, that are part of the translocon and by an increase of STT3B levels. On the one hand, the purpose of co-and posttranslational glycosylation of the different catalytic isoforms can be confirmed by these findings. On the other hand, they bring up the question of why STT3A and STT3B are not produced equally, to compensate for the loss of the each other subunit, and why shared members of the OST complex are going down to such an extend in the STT3A depleted cells.

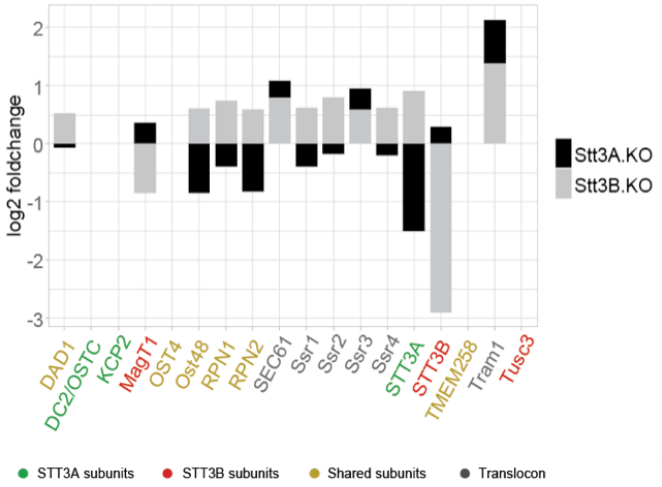


Figure 22: Translocon regulation of STT3A KO and STT3B KO including SEC61, TRAP complex, shared OST subunits and unique OST subunits

3.4 Pathway analysis

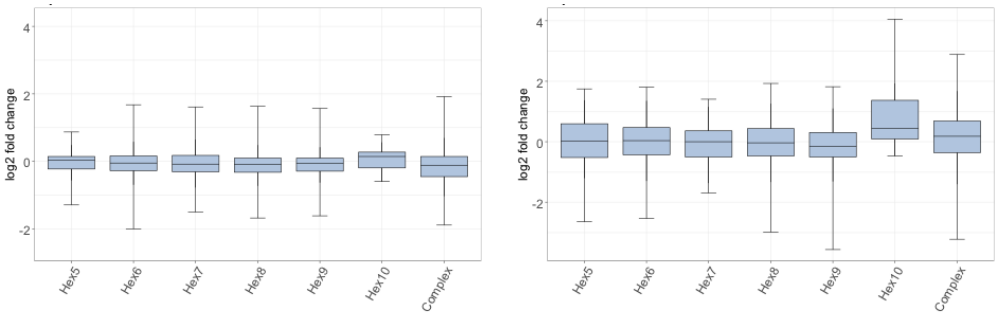
Gaining information on the impact on ER protein homeostasis and the cellular response after removal of one of the two distinct OST catalytic subunit isoforms needs evaluation of both the proteomic and glycoproteomic state of the different conditions. The analysis is focused on the UPR following ATF6, IRE1 and PERK pathways and their associated proteins as well as the ERAD.

	Gene name	STT3A - prot	STT3A - glyco	STT3B - prot	STT3B - glyco
ERAD	EDEM3	NA	-0.39	NA	-0.36
	CNX	+0.06	NA	+1.1	NA
	CRT	+0.4	NA	-0.3	NA
	HRD1	NA	NA	+0.69	NA
	Os9	NA	-0.1	NA	+0.68
	Uggt1	-0.25	NA	+0.19	NA
	Ganab	0.0	NA	-0.3	NA
	Sel1	-1.5	-0.43	+0.79	+0.65
ER chaperones and PDIs	GRP78	0.57	NA	-0.3	NA
	GRP94	-0.41	+2.2	-0.37	-0.28
	Pdia1	+0.38	-0.33	NA	NA
	Pdia3	+0.59	-1.16	-0.16	+0.5
	Pdia4	+0.44	-0.28	NA	NA
	Pdia6	+0.66	+0.33	NA	NA
	Dnajc10	0.0	NA	+0.60	NA
ATF6 UPR	Atf6	NA		NA	
PERK UPR	Eif2s1	0.0	NA	-0.52	NA
	Eif2s2	-0.26	NA	-0.38	NA
	Eif2s3	-0.28	NA	-0.73	NA
	Atf1	0.0	NA	1.62	NA
	Creb1	NA	NA	0.77	NA

Table 4: Identified UPR associated proteins in STT3A KO and STT3B KO cells

A closer investigation on UPR associated proteins for both STT3A and STT3B depleted cells did not show a clear indication, that UPR was forced following the well-described pathways of ATF6, IRE1 and PERK, in both conditions [Table 4]. GRP78, that is often used as stress induced UPR indicator, only showed a slight increase in the STT3A KO, whereas it appeared to be down regulated in STT3B KO cells. Moreover, no change in other ER derived chaperones could be observed, from one of these paths, but protein disulfate isomerases (Pdia 1-6) showed an increase in STT3A KO. However, eIF2 (Eif2s1-3), a translation initiation factor which is involved in the PERK mediated UPR, where the αsubunit gets phosphorylated, which then consequently results in decrease of translation, showed downregulation in the STT3B KO [117]. STT3A KO cells showed no change in the αsubunit of eIF2. Additionally, transcription factors Atf1 and Creb1, which are known to bind the GRP78 promotor in combination with Atf4, are present in much higher concentrations in the STT3B KO (no regulation in STT3A KO), although Atf4 could not be detected. According to literature, stress induced production increase of GRP78, using tunicamycin for instance, critically depends on the complex formation of all three transcription factors [118]. In contrast to these finding ERAD, as part of UPR, seems to be upregulated in the STT3B KO, when comparing ERAD associated proteins to native and STT3A KO conditions. Glycan distribution analysis revealed an upregulation of Hex10 glycan structure containing peptides in STT3B depleted cells, compared to those with complex and low-mannose glycan structures attached [Fig. 23]. Presence of high-mannose structures with termi-

nal glucose residues like the Hex10 structure indicates an upregulation of the calnexin/calreticulin cycle, although the amount of concerned peptides still is low [S 5].



(a) STT3A KO vs. WT (b) STT3B KO vs. WT

Figure 23: Regulation of peptides containing different glycan species at PSM level in STT3A KO cells and STT3B KO cells.

Moreover, downregulation of translation factors is a large branch of UPR. Reducing the amount of newly synthesized proteins over some time, is used to restore cell homeostasis and prevent apoptosis. Bioinformatic BP-analysis using DAVID and STRING showed a decrease of translation associated proteins both in STT3A KO and STT3B KO cells, indicating partial UPR response in both cases. Ribosomal proteins (Rpl18, Rpl4, Rpl7, Rpl19, Rpl30, Rps13, Rps3, Rps23, Rps29) as well as translation initiation (Eif3a-m) and tRNA synthesis associated proteins (Aars, Gars, Iars) resulted in a high reduction in STT3B depleted cells [S 11]. Also, a decrease in ribosomal proteins, translation initiation proteins, translation elongation factors (Eef1) and rRNA processing proteins could be observed in STT3A KO cells [S 7]. Knockout of each catalytic subunits of both OST isoforms seems to activate translational part of UPR, ERAD only seems to be increased upon STT3B knockout.

Analysis of significantly upregulated proteins (p=0.05) showed an increase of proteins related to ER stress response and protein folding (Hyou1, Eef2, Hspa5, P4hb, Pdia1-6, Cct4) and respiratory (electron) transport chain (Hibadh, Ndufa1-7, Ndufb1-6, Ndufv2,3, Fasn, Sdhb, Uqcrh) in the STT3A KO [S 6]. STT3B KO led to upregulation of proteins associated to nucleosome assembly and transcription regulation [S 10]. Investigation of the general expression behavior among peptides and glycosylated peptides displayed a larger influence of knockout of the STT3B subunit on protein regulation. Comparison of shared peptides among both KO states indicated a stronger dependence on STT3B, than STT3A [Fig. 24]. NxT/S/C analysis of the glycoproteomics data set also showed a dependence, that is enhanced in the STT3B KO. As control, potential non-annotated glycosylation sequons were analyzed in the proteomics dataset [S 14, 15].

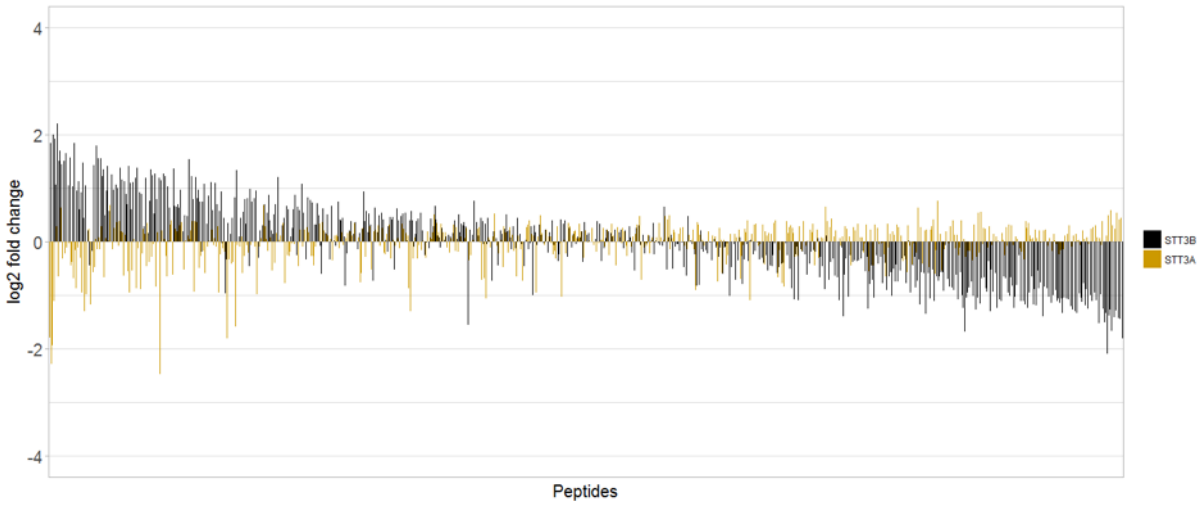
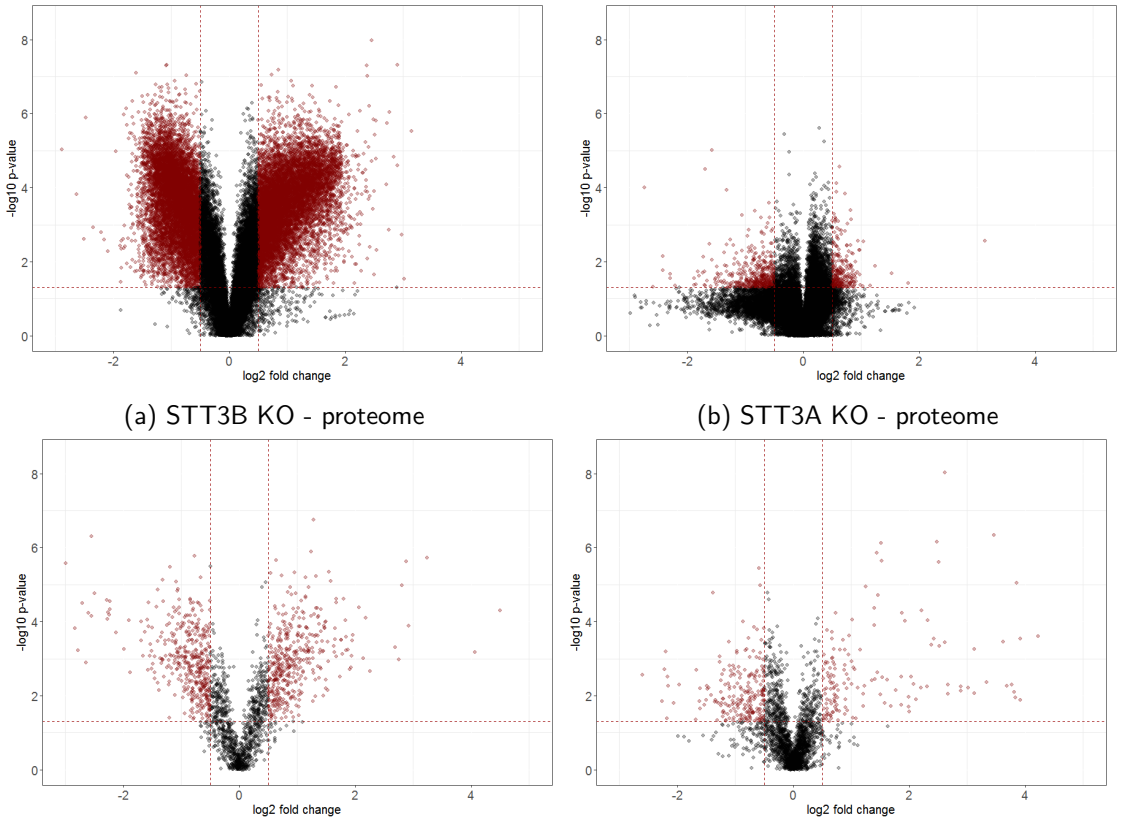


Figure 24: Comparison of regulation among shared peptides between STT3A KO and STT3B KO



(a) STT3B KO - proteome

(b) STT3A KO - proteome

(c) STT3B KO - glycoproteome

(d) STT3A KO - glycoproteome

Figure 25: Volcano plot of a) proteomics data of STT3B KO vs. WT cells at PSM levels, b) proteomics data of STTA KO vs. WT cells at PSM levels, c) glycoproteomics data of STT3B KO vs. WT cells at PSM levels, d) glycoproteomics data of STT3A KO vs. WT cells at PSM levels.

Moreover, STT3B KO led to a general strong protein response when looking at upregulated and downregulated PSM levels, which were evenly distributed. Glycoprotein PSMs showed a similar response of STT3B KO, where glycosylation sites were equally affected by up- and downregulation [Fig. 25a, c]. STT3A kockout resulted in a higher amount of significantly upregulated proteins PSMs ($p<0.05$) and low fold changes. Stronger and negative regulation effects of STT3A KO could be seen in a smaller amount of downregulated protein PSM. A higher proportion of glycoproteins, which showed a positive foldchange, were affected by STT3A KO. Upregulation of glycoprotein PSM hence indicated either upregulation of the protein or the glycosylation site [Fig. 25b, d].

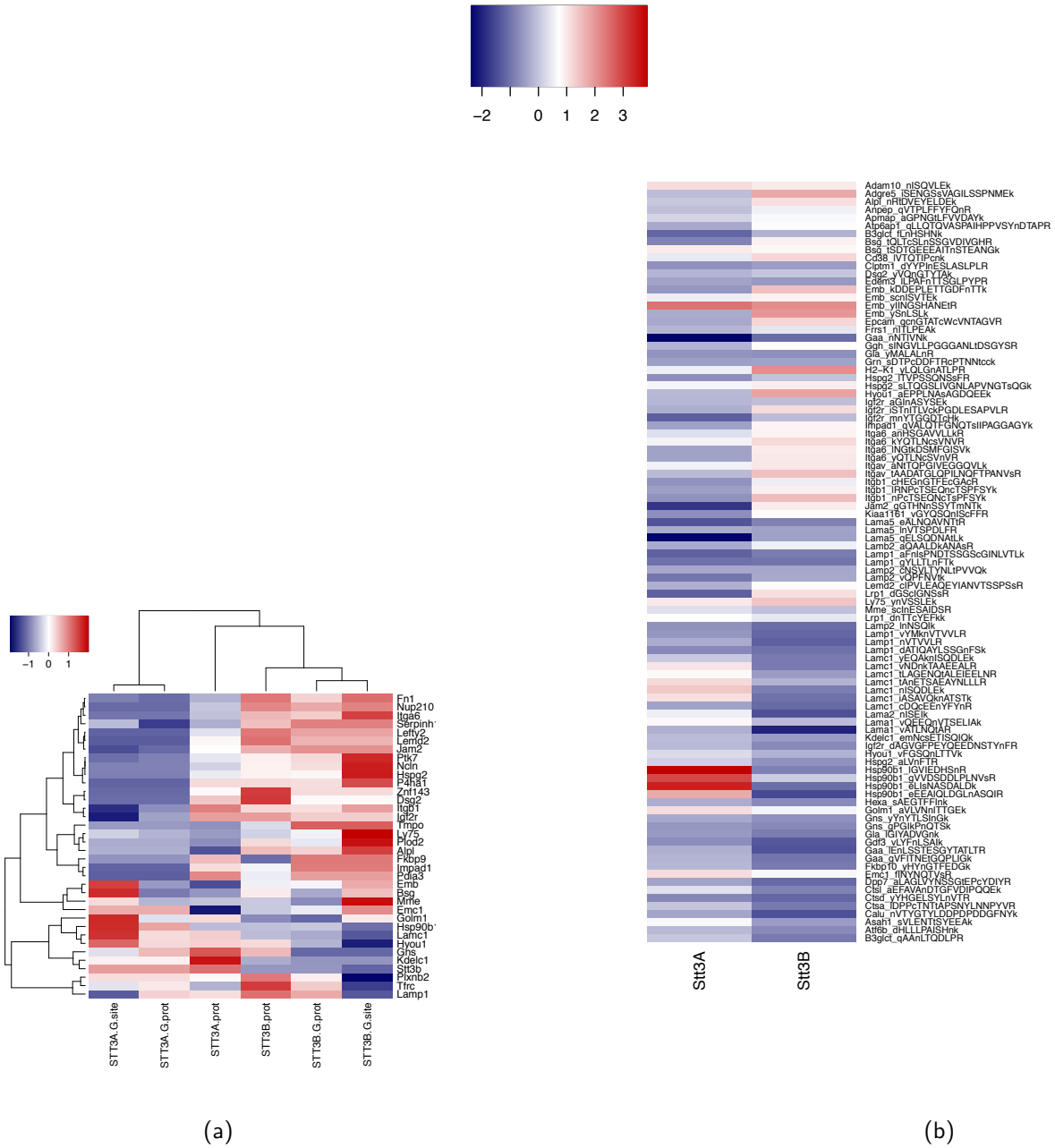
A small group of 36 proteins, found in both hpH- and HILIC-enriched samples as well as in STT3A KO and STT3B KO cells was compared by their translation levels [Fig. 26a]. Protein abundance was correlated with the according glycosylated protein variant, both for global glycosylation and "extreme"-glycosylated sequons. Whereas most proteins showed a regulatory effect, influencing glycosylation and protein translation in the same way, several proteins[1] where found, which showed a contrary glycosylation efficiency to their translational regulation.

Genes: P4ha1, Dsg2, Itgb2, Igf2r, Fkbp9, Impad1, Pdia3, Emb, Bsg, MMe, Emc1, Golm1, Hsp90b, Gns, Tfrc, Lamp1, Tmpo, Ly75, Hyou1, Plxnb2 [1]

This group of proteins got reduced to the number of proteins[2] that are not influenced by STT3A and STT3B in the same way.

Genes: P4ha1, Dsg2, Itgb2, Igf2r, Fkbp9, Impad1, Pdia3, Golm1, Hsp90b, Tfrc, Tmpo, Ly75, Hyou1, Plxnb2 [2]

Moreover, only few proteins showed site specific glycosylation regulation, that contradicts global protein glycosylation and translation. Comparison of glycopeptides, that are shared between STT3A KO and STT3B KO, again showed a larger amount of glycosylation sites, that are stronger influenced by STT3B [Fig. 26b]. Whereas most glycopeptides regulated in the same direction, a small population was upregulated in the STT3A KO and a larger population was upregulated in the STT3B KO. Clustering of the glycopeptides indicated a trend of influence on glycosylation for both knockout samples.



(a) (b)

Figure 26: a) Clustering of shared glycoproteins among both KO samples, that could be recovered in the proteomics dataset for comparison. The mean value for all PSMs of either a peptide (STT3A/B.G.site) or a protein (STT3A/B.(G.)prot) was used. b) Clustering of shared peptides among both KO samples to evaluate sequence specificity of the catalytical subunits.

4 Discussion

Chromatographic approaches operated at ph levels higher than the traditionally used acidic pH gradients, have proven to be robust and reproducible systems. In this study, the peptide separation efficiency as well as the capability to decomplex the sample, could be evidenced by the high recovery between STT3A KO and STT3B KO samples. However, heterogeneity in different glycoforms is still a limiting factor in glycoproteomics studies. It complicates the development of a truly homogenous applicable method, that is able to cover peptides carrying this manifold modification, while still keeping chromatographic and mass spectrometric measuring time in a realistic range. Moreover, glycoproteins are often present in very low abundance and located in membranes to a large extend. Technical errors are therefore possible to occur yet in the early stages of sample preparation, during cell lysis and digestion. The reduced recovery between the samples can be explained by this fact, as well as by the maintenance-susceptibly HILIC system, that needs long equilibration times and is prone to contamination. Nevertheless, in this study, HILIC is used as a compromise between an optimized detection of a large population of different intact glycopeptides and more time-consuming specialized enrichment strategies. Reproducibility between runs, however, should be further tested and optimized.

The large discrepancy between expression and translation levels can be visualized by comparing proteomics with RNA-seq data. Especially for glycoproteomics studies aiming for comparative sequon glycosylation efficiencies, the limitation of measuring low abundant glycoproteins in a proteomics approach is decisive. Pulse-chase analysis is widely used for detection of N-glycosylation on proteins and can be used comparatively [119, 120]. However, proteome wide studies that exceed substrate directed analysis demand for a more general approach. Up to now, mass spectrometry is the only method to analyze intact glycoproteins at high throughput.

Pulse-chase technology has been used in experiments verifying site specificity of STT3B [119]. These studies are mainly based on insertion of expression vectors, containing glycoprotein substrates, in host cells. RNAi knockdowns as well as CRISPR knockouts have been used to remove either the post-translationally or co-translationally modifying catalytic subunit of the OST complex [119]. In contrast to the present study, HEK or HeLa cells are mainly used at comparably high glucose concentration levels (5%). Although STT3A has been determined as main factor in protein glycosylation and according to literature is essential for cell survival since its knockout induces UPR, sequence specificities have only been found for STT3B [119]. Sterical hindrance is postulated as problem, that seems to interfere with STT3A's catalytical activity on adjacent sequons. In contrast, closely spaced acceptor sites, that are quite common in glycoproteins according to sequon analysis on combined murine protein database, may very well be glycosylated by STT3A [14]. To review this hypothesis, mESC glycoproteomics data from STT3A KO and STT3B KO was filtered for peptides containing more than one NxT/S/C sequon per sequence. Only four peptides could be identified, that carrying motifs, that can be described as adjacent or closely spaced [Table 4]. Additionally, the peptides showed no contrary regulation between STT3A and STT3B KO. In case of Lamp1 and Emb, the specific

glycosylation site does not represent the overall glycosylation behavior upon knockout of one of the subunits. Hence, their closely located acceptor sites seem to be dependent on the interplay of both subunits, but not on a single one of them. Influence on glycosylation efficiency of the co-translational subunit due to sterical effects have also been observed on sequons containing bulky hydrophobic or negatively charged amino acids such as W, L, F, I, M and D [15]. According to this theory, STT3B KO should induce reduction of peptides, containing glycosylation sites with such amino acids on middle position. In this study, affected downregulated peptides in the STT3B KO showed a relative percentage being higher than the one of upregulated peptides for L, M, F and D. Still, this effect is balanced by the absolute quantity of peptides carrying this motif, which is higher in the pool of upregulated peptides. Comparison of this two situations would demand an equal set of peptides, which in this case differ by n=47. Therefore, no conclusion on a sequence dependency of STT3B can be made.

Glycosylation of sequons close to the C-terminus or close to signal peptides at the N-terminus of secretory proteins has been found to rely on the post translational glycosylation apparatus as well [12, 13]. A precondition to evaluate this dependency is a significant amount of glycopeptides that fit this location. In prior studies, vector inserted model glycoproteins, showing C-terminal glycosites have been used in HeLa cells displaying STT3B influence [12]. A bioinformatics approach to screen glycoproteomics data for glycosylation efficiency both upon STT3A and STT3B knockout is obligatory but has been missing so far. However, the typical distribution of glycosylation sites over a protein is not favorable for terminal glycosylation. Analysis of sequons shows a strong centralization of acceptor sites, whereas C-terminal domains as well as N-terminal domains were hardly glycosylated. By definition, C-terminal glycosites should lie within a range of 75 amino acid residues from the terminus. Filtering of glycopeptides with sites located closer than 100 residues away from the terminus clearly shows an under-representation of such glycosites. At PSM level a slightly higher amount of negatively regulated glycosites is visible in STT3B KO data, compared to STT3A KO [Fig. 21b, c]. Yet, there is a general trend at glycol-PSM level to be downregulated as shown in S15. Moreover, proteins with C-terminal glycosites showed a glycosylation with similar trend for different peptides. Acceptor sites close to a signal peptide are concerned in the same way by the general glycosylation distribution. Few peptides where found with N-terminal glycosylation. Those which contained a signal peptide sequence showed no downregulation of glycosylation in the STT3B KO. No sequence dependency on STT3B of terminal glycosites could be therefore determined for this experiment using mESCs.

Since STT3A is thought to be responsible for the bulk glycosylation in the cell, a full KO is expected to cause major stress and massively alter glycosylation behavior of the cells. RNA-seq data confirms the high demand of the cell on STT3A (FPKM = 147) with expression levels more than three times higher than those of STT3B (FPKM = 44). Cells, such as mESC in this study, lacking STT3A or STT3B are still capable of glycosylating proteins to a large extend, with no dominant restriction that leads to an impaired cell viability. One of the main questions is, how the glycosylation machinery can compensate for a loss of its major component. In a former study, knockout of catalytic subunit isoforms in the OST complex has been found studies to influence translation of

other members of the OST and translocon complex as well [119]. Most prominently, data of the current study shows the compensatory effect of the each other catalytic subunit. However, while STT3B KO induces high upregulation of STT3A, STT3B is only slightly upregulated in the STT3A KO. Additionally, shared members of the OST complex, as well as the translocon members act in an exclusively compensatory way in the STT3B knockout. Translation levels of most shared OST subunits are going down in the STT3A KO, which once again points out the higher expression of STT3A in mECSs. Apparently, the main part of the OST complex is present translocon associated in combination with STT3A.

It is important to note, that whereas both knockouts produce viable clones, a full knockout of both subunits leads to lethality [13]. However, removing one of the key elements of the ER glycosylation machinery, which is further necessary to maintain protein homeostasis in the ER, due to its participation in protein folding and degradation of misfolded proteins, is expected to have a major influence on glycosylation levels and further on protein processing in the ER. Tunicamycin derived inhibition of N-glycosylation is often used to induce ER stress and has been found to initiate unfolded protein response and ER stress can be activated by targeting the co-translational glycosylating subunit [13, 121]. Under ER stress conditions UPR acts as regulatory machine to restore cell homeostasis and prevent apoptosis by firstly reducing translation and activating ERAD and secondly by increasing production of chaperones to improve protein folding. UPR response is generally driven by the activation of three different pathways following ATF6, IRE1 and PERK, which are further necessary to activate ERAD associated genes, to active increased chaperone production, to stop protein translation or to induce apoptosis and autophagy [117, 122, 123]. Detection of these pathway initiating proteins, ERAD associated proteins (EDEM1-3, CNX/CRT, HRD1, OS9, HERP) as well as cell stress responsive chaperones such as BiP/GRP78 and GRP94 can be used to monitor ER stress induced UPR [124]. ATF6, XBP1s, p-eIF2, GRP78, GRP94 and MANF for instance were used to detect UPR response in the mouse brain among tunicamycin induced ER stress [13,121].

Here, proteomics analysis of mESCs showed no upregulated UPR in STT3A- and a slightly upregulated UPR in STT3B KO cells [Table 4]. The key pathways following PERK, IRE1 and Atf6 showed a marginal regulation. Only STT3B KO resulted in a regulation of PERK associated proteins. Unexpectedly, only knockout of STT3B led to an upregulation of the ERAD. In contrast, STT3A KO experiences downregulation of Sel1, one of the key player in the ERAD machinery. Sel1 is responsible for transportation of misfolded proteins to the cytoplasm for protein degradation. Moreover, GRP94 gets highly glycosylated in the STT3A KO. GRP94 has been found to specifically play a role in protein removal from the ER by interacting with Sel1 [18]. While explaining the reduced ERAD functionality, the question comes up, whether it is caused by a general loss of glycosylated sites, being necessary for protein folding, or by ERAD associated proteins. Many ER residential proteins, including ERAD proteins, are glycoproteins themselves with lectin functions. Their glycosylation may be STT3A dependent. The impaired ERAD of STT3A depleted cells is mirrored by the downregulation of peptides, that contain complex or hybrid glycan structures [Fig. 23]. Due to improper folding, less glycoproteins can leave the ER for further glycan processing in the Golgi. Although translational

regulation is reduced in both knockouts and ERAD is upregulated in STT3B KO, the expected effect of a high UPR response is not visible.

The unexpectedly higher regulation of proteins and glycoproteins upon STT3B KO interferes with the idea of STT3B only being responsible for missing glycosites. Knockout of STT3B shows a higher effect on glycosites and more glycoproteins are concerned. Still, a site specificity is not visible, neither for STT3A nor for STT3B. In any case, protein related dependencies can be seen when clustering shared glycopeptides. Most peptides of the same protein are affected in the same way by glycosylation, for both subunits of the OST complex. Clearly, further investigation on the OST complex and its potential substrate and sequon dependencies is necessary. Obligatory is the comparison of different model systems and cell lines, that contain high glycosylation activity. Methodological improvement to increase the depth of analysis for glycopeptides is necessary but still challenging. The usage of a quantitative intact glycoproteomics approach could be very helpful to follow glycoprotein distribution and to answer questions on glycosylation conditions of proteins and cells.

References

[1] Sean R. Stowell, Tongzhong Ju, and Richard D. Cummings. Protein glycosylation in cancer. *Annual Review of Pathology: Mechanisms of Disease*, 10(1):473–510, Jan 2015.

[2] Kelley W. Moremen, Michael Tiemeyer, and Alison V. Nairn. Vertebrate protein glycosylation: diversity, synthesis and function. *Nature Reviews Molecular Cell Biology*, 13(7):448–462, Jul 2012.

[3] Ajit Varki. *Essentials of glycobiology*. Third edition edition.

[4] Jamey D Marth and Prabhjit K Grewal. Mammalian glycosylation in immunity. *Nat Rev Immunol*, 8(11):874–87, Nov 2008.

[5] Margreet A Wolfert and Geert-Jan Boons. Adaptive immune activation: glycosylation does matter. *Nat Chem Biol*, 9(12):776–84, Dec 2013.

[6] Salomé S. Pinho and Celso A. Reis. Glycosylation in cancer: mechanisms and clinical implications. *Nature Reviews Cancer*, 15(9):540–555, Aug 2015.

[7] Stephanie Grunewald, Gert Matthijs, and Jaak Jaeken. Congenital disorders of glycosylation: a review. *Pediatr Res*, 52(5):618–24, Nov 2002.

[8] Jaak Jaeken and Gert Matthijs. Congenital disorders of glycosylation. *Annual Review of Genomics and Human Genetics*, 2(1):129–151, 2001. PMID: 11701646.

[9] Jaak Jaeken. Congenital disorders of glycosylation. *Annals of the New York Academy of Sciences*, 1214(1):190–198, Dec 2010.

[10] Hudson H Freeze, Erik A Eklund, Bobby G Ng, and Marc C Patterson. Neurology of inherited glycosylation disorders. *Lancet Neurol*, 11(5):453–66, May 2012.

[11] Katharina Braunger, Stefan Pfeffer, Shiteshu Shrimal, Reid Gilmore, Otto Berninghausen, Elisabet C Mandon, Thomas Becker, Friedrich Förster, and Roland Beckmann. Structural basis for coupling protein transport and n-glycosylation at the mammalian endoplasmic reticulum. *Science*, 360(6385):215–219, 04 2018.

[12] Shiteshu Shrimal, Steven F. Trueman, and Reid Gilmore. Extreme c-terminal sites are posttranslocationally glycosylated by the stt3b isoform of the ost. *The Journal of Cell Biology*, 201(1):81–95, Mar 2013.

[13] Catalina Ruiz-Canada, Daniel J. Kelleher, and Reid Gilmore. Cotranslational and posttranslational n-glycosylation of polypeptides by distinct mammalian ost isoforms. *Cell*, 136(2):272–283, Jan 2009.

[14] S. Shrimal and R. Gilmore. Glycosylation of closely spaced acceptor sites in human glycoproteins. *Journal of Cell Science*, 126(23):5513–5523, Oct 2013.

[15] Heidi L H Malaby and William R Kobertz. The middle x residue influences cotranslational n-glycosylation consensus site skipping. *Biochemistry*, 53(30):4884–93, Aug 2014.

[16] Birgit Meusser, Christian Hirsch, Ernst Jarosch, and Thomas Sommer. Erad: the long road to destruction. *Nat Cell Biol*, 7(8):766–72, Aug 2005.

[17] John C. Christianson, Thomas A. Shaler, Ryan E. Tyler, and Ron R. Kopito. Os-9 and grp94 deliver mutant Îś1-antitrypsin to the hrd1–sel1l ubiquitin ligase complex for erad. *Nature Cell Biology*, 10(3):272–282, Feb 2008.

[18] James A Olzmann, Ron R Kopito, and John C Christianson. The mammalian endoplasmic reticulum-associated degradation system. *Cold Spring Harb Perspect Biol*, 5(9), Sep 2013.

[19] Karen L. Abbott and J. Michael Pierce. Lectin-based glycoproteomic techniques for the enrichment and identification of potential biomarkers. *Glycobiology*, pages 461–476, 2010.

[20] Yuan Tian, Kay Gurley, Danni L Meany, Christopher J Kemp, and Hui Zhang. N-linked glycoproteomic analysis of formalin-fixed and paraffin-embedded tissues. *J Proteome Res*, 8(4):1657–62, Apr 2009.

[21] E K Fridriksson, B Baird, and F W McLafferty. Electrospray mass spectra from protein electroeluted from sodium dodecylsulfate polyacrylamide gel electrophoresis gels. *J Am Soc Mass Spectrom*, 10(5):453–5, May 1999.

[22] Yee-Guide Yeung, Edward Nieves, Ruth Hogue Angeletti, and E Richard Stanley. Removal of detergents from protein digests for mass spectrometry analysis. *Anal Biochem*, 382(2):135–7, Nov 2008.

[23] Dorota F. Zielinska, Florian Gnad, Jacek R. Wiśniewski, and Matthias Mann. Precision mapping of an in vivo n-glycoproteome reveals rigid topological and sequence constraints. *Cell*, 141(5):897–907, May 2010.

[24] Sheng Pan, Ru Chen, Ruedi Aebersold, and Teresa A. Brentnall. Mass spectrometry based glycoproteomics—from a proteomics perspective. *Molecular Cellular Proteomics*, 10(1):R110.003251, Aug 2010.

[25] H H Freeze. Lectin affinity chromatography. *Curr Protoc Protein Sci*, Chapter 9:Unit 9.1, May 2001.

[26] Susan Fanayan, Marina Hincapie, and William S. Hancock. Using lectins to harvest the plasma/serum glycoproteome. *ELECTROPHORESIS*, 33(12):1746–1754, Jun 2012.

[27] Eliel Ruiz-May, Carmen Catalá, and Jocelyn K C Rose. N-glycoprotein enrichment by lectin affinity chromatography. *Methods Mol Biol*, 1072:633–43, 2014.

[28] R D Cummings and S Kornfeld. Fractionation of asparagine-linked oligosaccharides by serial lectin-agarose affinity chromatography. a rapid, sensitive, and specific technique. *J Biol Chem*, 257(19):11235–40, Oct 1982.

[29] Malaika Durham and Fred E Regnier. Targeted glycoproteomics: serial lectin affinity chromatography in the selection of o-glycosylation sites on proteins from the human blood proteome. *J Chromatogr A*, 1132(1-2):165–73, Nov 2006.

[30] Meenakshi Dowlut and Dennis G Hall. An improved class of sugar-binding boronic acids, soluble and capable of complexing glycosides in neutral water. *J Am Chem Soc*, 128(13):4226–7, Apr 2006.

[31] Ying Zhang, Jing Jiao, Pengyuan Yang, and Haojie Lu. Mass spectrometry-based n-glycoproteomics for cancer biomarker discovery. *Clin Proteomics*, 11(1):18, 2014.

[32] Haopeng Xiao, Weixuan Chen, Johanna M. Smeekens, and Ronghu Wu. An enrichment method based on synergistic and reversible covalent interactions for large-scale analysis of glycoproteins. *Nature Communications*, 9(1), Apr 2018.

[33] Martin R. Larsen, Søren S. Jensen, Lene A. Jakobsen, and Niels H. H. Heegaard. Exploring the sialiome using titanium dioxide chromatography and mass spectrometry. *Molecular Cellular Proteomics*, 6(10):1778–1787, Jul 2007.

[34] Kasper Engholm-Keller and Martin R. Larsen. Titanium dioxide as chemo-affinity chromatographic sorbent of biomolecular compounds — applications in acidic modification-specific proteomics. *Journal of Proteomics*, 75(2):317–328, Dec 2011.

[35] Yingwei Hu, Punit Shah, David J. Clark, Minghui Ao, and Hui Zhang. Reanalysis of global proteomic and phosphoproteomic data identified a large number of glycopeptides. *Analytical Chemistry*, 90(13):8065–8071, May 2018.

[36] Giuseppe Palmisano, Sara Eun Lendal, Kasper Engholm-Keller, Rikke Leth-Larsen, Benjamin L Parker, and Martin R Larsen. Selective enrichment of sialic acid–containing glycopeptides using titanium dioxide chromatography with analysis by hilic and mass spectrometry. *Nature Protocols*, 5(12):1974–1982, Nov 2010.

[37] Hui Zhang, Xiao-Jun Li, Daniel B Martin, and Ruedi Aebersold. Identification and quantification of n-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat Biotechnol*, 21(6):660–6, Jun 2003.

[38] Bogusław Buszewski and Sylwia Noga. Hydrophilic interaction liquid chromatography (hilic)—a powerful separation technique. *Analytical and Bioanalytical Chemistry*, 402(1):231–247, Aug 2011.

[39] Andrew J. Alpert. Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. *Journal of Chromatography A*, 499:177 – 196, 1990.

[40] Tatsunari Yoshida. Peptide separation in normal phase liquid chromatography. *Analytical Chemistry*, 69(15):3038–3043, 08 1997.

[41] Serena Di Palma, Shabaz Mohammed, and Albert J R Heck. Zic-chilic as a fractionation method for sensitive and powerful shotgun proteomics. *Nature Protocols*, 7(11):2041–2055, Oct 2012.

[42] Per Hägglund, Torny Eriksson, Anna Collén, Wim Nerinckx, Marc Claeyssens, and Henrik Stålbrand. A cellulose-binding module of the trichoderma reesei beta-mannanase man5a increases the mannan-hydrolysis of complex substrates. *J Biotechnol*, 101(1):37–48, Feb 2003.

[43] Paul J. Boersema, Shabaz Mohammed, and Albert J. R. Heck. Hydrophilic interaction liquid chromatography (hilic) in proteomics. *Analytical and Bioanalytical Chemistry*, 391(1):151–159, Feb 2008.

[44] Petrus Hemström and Knut Irgum. Hydrophilic interaction chromatography. *Journal of Separation Science*, 29(12):1784–1821, Aug 2006.

[45] Kathirvel Alagesan, Sana Khan Khilji, and Daniel Kolarich. It is all about the solvent: on the importance of the mobile phase for zic-hilic glycopeptide enrichment. *Analytical and Bioanalytical Chemistry*, 409(2):529–538, Dec 2016.

[46] Gerhild Zauner, André M. Deelder, and Manfred Wuhrer. Recent advances in hydrophilic interaction liquid chromatography (hilic) for structural glycomics. *ELECTROPHORESIS*, 32(24):3456–3466, Dec 2011.

[47] Simon Mysling, Giuseppe Palmisano, Peter Højrup, and Morten Thaysen-Andersen. Utilizing ion-pairing hydrophilic interaction chromatography solid phase extraction for efficient glycopeptide enrichment in glycoproteomics. *Anal Chem*, 82(13):5598–609, Jul 2010.

[48] Dalila Bensaddek, Armel Nicolas, and Angus I Lamond. Evaluating the use of hilic in large-scale, multi dimensional proteomics: Horses for courses? *Int J Mass Spectrom*, 391:105–114, Nov 2015.

[49] John R Yates III. A century of mass spectrometry: from atoms to proteomes. *Nature Methods*, 8:633 EP –, 07 2011.

[50] Ruedi Aebersold and Matthias Mann. Mass-spectrometric exploration of proteome structure and function. *Nature*, 537(7620):347–355, Sep 2016.

[51] Tiannan Guo, Petri Kouvonen, Ching Chiek Koh, Ludovic C Gillet, Witold E Wolski, Hannes L Röst, George Rosenberger, Ben C Collins, Lorenz C Blum, Silke Gillessen, and et al. Rapid mass spectrometric conversion of tissue biopsy samples into permanent quantitative digital proteome maps. *Nature Medicine*, 21(4):407–413, Mar 2015.

[52] Piero Giansanti, Liana Tsiatsiani, Teck Yew Low, and Albert J R Heck. Six alternative proteases for mass spectrometry–based proteomics beyond trypsin. *Nature Protocols*, 11(5):993–1006, Apr 2016.

[53] Vladimir G Zaikin and John M Halket. Derivatization in mass spectrometry–8. soft ionization mass spectrometry of small molecules. *Eur J Mass Spectrom (Chichester)*, 12(2):79–115, 2006.

[54] Jens S. Andersen, Birte Svensson, and Peter Roepstorff. Electrospray ionization and matrix assisted laser desorption/ionization mass spectrometry: Powerful analytical tools in recombinant protein chemistry. *Nature Biotechnology*, 14:449 EP –, 04 1996.

[55] Victor V. Laiko, Michael A. Baldwin, and Alma L. Burlingame. Atmospheric pressure matrix-assisted laser desorption/ionization mass spectrometry. *Analytical Chemistry*, 72(4):652–657, 2000. PMID: 10701247.

[56] Gary L. Glish and Richard W. Vachet. The basics of mass spectrometry in the twenty-first century. *Nature Reviews Drug Discovery*, 2(2):140–150, Feb 2003.

[57] Thorsten W. Jaskolla and Michael Karas. Compelling evidence for lucky survivor and gas phase protonation: The unified maldi analyte protonation mechanism. *Journal of The American Society for Mass Spectrometry*, 22(6):976–988, Mar 2011.

[58] M Karas, M Glückmann, and J Schäfer. Ionization in matrix-assisted laser desorption/ionization: singly charged molecular ions are the lucky survivors. *J Mass Spectrom*, 35(1):1–12, Jan 2000.

[59] Richard Zenobi, Renato Knochenmuss. Ion formation in maldi mass spectrometry. *Mass Spectrometry Reviews*, 17, 1998.

[60] JB Fenn, M Mann, CK Meng, SF Wong, and CM Whitehouse. Electrospray ionization for mass spectrometry of large biomolecules. *Science*, 246(4926):64–71, 1989.

[61] Disintegration of water drops in an electric field. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 280(1382):383–397, 1964.

[62] Matthias Wilm. Principles of electrospray ionization. *Molecular Cellular Proteomics*, 10(7):M111.009407, Jul 2011.

[63] Klaus Biemann. [25] sequencing of peptides by tandem mass spectrometry and high-energy collision-induced dissociation. In *Mass Spectrometry*, volume 193 of *Methods in Enzymology*, pages 455 – 479. Academic Press, 1990.

[64] J Mitchell Wells and Scott A McLuckey. Collision-induced dissociation (cid) of peptides and proteins. *Methods Enzymol*, 402:148–85, 2005.

[65] Andrew W. Jones and Helen J. Cooper. Dissociation techniques in mass spectrometry-based proteomics. *The Analyst*, 136(17):3419, 2011.

[66] Yu Xia, Xiaorong Liang, and Scott A McLuckey. Ion trap versus low-energy beam-type collision-induced dissociation of protonated ubiquitin ions. *Anal Chem*, 78(4):1218–27, Feb 2006.

[67] Jesper V Olsen, Boris Macek, Oliver Lange, Alexander Makarov, Stevan Horning, and Matthias Mann. Higher-energy c-trap dissociation for peptide modification analysis. *Nat Methods*, 4(9):709–12, Sep 2007.

[68] Mark P. Jedrychowski, Edward L. Huttlin, Wilhelm Haas, Mathew E. Sowa, Ramin Rad, and Steven P. Gygi. Evaluation of hcd- and cid-type fragmentation within their respective detection platforms for murine phosphoproteomics. *Molecular Cellular Proteomics*, 10(12):M111.009910, Sep 2011.

[69] J. E. P. Syka, J. J. Coon, M. J. Schroeder, J. Shabanowitz, and D. F. Hunt. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proceedings of the National Academy of Sciences*, 101(26):9528–9533, Jun 2004.

[70] Julia Wiesner, Thomas Premsler, and Albert Sickmann. Application of electron transfer dissociation (etd) for the analysis of posttranslational modifications. *PROTEOMICS*, 8(21):4466–4483, Nov 2008.

[71] David M. Good, Matthew Wirtala, Graeme C. McAlister, and Joshua J. Coon. Performance characteristics of electron transfer dissociation mass spectrometry. *Molecular Cellular Proteomics*, 6(11):1942–1951, Aug 2007.

[72] Nicholas M. Riley, Michael S. Westphall, and Joshua J. Coon. Activated ion electron transfer dissociation for improved fragmentation of intact proteins. *Analytical Chemistry*, 87(14):7109–7116, Jun 2015.

[73] Danielle L Swaney, Graeme C McAlister, Matthew Wirtala, Jae C Schwartz, John E P Syka, and Joshua J Coon. Supplemental activation method for high-efficiency electron-transfer dissociation of doubly protonated peptide precursors. *Anal Chem*, 79(2):477–85, Jan 2007.

[74] Michael A Gillette and Steven A Carr. Quantitative analysis of peptides and proteins in biomedicine by targeted mass spectrometry. *Nature Methods*, 10(1):28–34, Jan 2013.

[75] Hannes L Röst, George Rosenberger, Pedro Navarro, Ludovic Gillet, Saša M Miladinović, Olga T Schubert, Witold Wolski, Ben C Collins, Johan Malmström, Lars Malmström, and Ruedi Aebersold. Openswath enables automated, targeted analysis of data-independent acquisition ms data. *Nature Biotechnology*, 32:219 EP –, 03 2014.

[76] Florian Meier, Philipp E Geyer, Sebastian Virreira Winter, Juergen Cox, and Matthias Mann. Boxcar acquisition method enables single-shot proteomics at a depth of 10,000 proteins in 100 minutes. *Nat Methods*, May 2018.

[77] Allison Doerr. Dia mass spectrometry. *Nature Methods*, 12:35 EP –, 12 2014.

[78] R J Cotter. Peer reviewed: The new time-of-flight mass spectrometry. *Anal Chem*, 71(13):445A–51A, Jul 1999.

[79] W C. Wiley and I H. McLaren. TimeâĂŘofâĂŘflight mass spectrometer with improved resolution. *Review of Scientific Instruments*, 26:1150 – 1157, 01 1956.

[80] B.A Mamyrin. Time-of-flight mass spectrometry (concepts, achievements, and prospects). *International Journal of Mass Spectrometry.*, 206:251–266, 2001.

[81] Igor V. Chernushevich, Alexander V. Loboda, and Bruce A. Thomson. An introduction to quadrupole–time-of-flight mass spectrometry. *Journal of Mass Spectrometry*, 36(8):849–865.

[82] Wolfgang Paul. Electromagnetic traps for charged and neutral particles. *Review of modern physics*, 62(3):513, 1990.

[83] Adele R Blackler, Aaron A Klammer, Michael J MacCoss, and Christine C Wu. Quantitative comparison of proteomic data quality between a 2d and 3d quadrupole ion trap. *Anal Chem*, 78(4):1337–44, Feb 2006.

[84] R. A. Yost and C. G. Enke. Selected ion fragmentation with a tandem quadrupole mass spectrometer. *Journal of the American Chemical Society*, 100(7):2274–2275, 1978.

[85] Alessio Maiolica, Martin A. Jünger, Iakes Ezkurdia, and Ruedi Aebersold. Targeted proteome investigation via selected reaction monitoring mass spectrometry. *Journal of Proteomics*, 75(12):3495–3513, Jun 2012.

[86] Paola Picotti and Ruedi Aebersold. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat Methods*, 9(6):555–66, May 2012.

[87] Silvia Surinova, Ruth Hüttenhain, Ching-Yun Chang, Lucia Espona, Olga Vitek, and Ruedi Aebersold. Automated selected reaction monitoring data analysis workflow for large-scale targeted proteomic studies. *Nat Protoc*, 8(8):1602–19, Aug 2013.

[88] Michael A Kuzyk, Derek Smith, Juncong Yang, Tyra J Cross, Angela M Jackson, Darryl B Hardie, N Leigh Anderson, and Christoph H Borchers. Multiple reaction monitoring-based, multiplexed, absolute quantitation of 45 proteins in human plasma. *Mol Cell Proteomics*, 8(8):1860–77, Aug 2009.

[89] Alexander Makarov, Eduard Denisov, Alexander Kholomeev, Wilko Balschun, Oliver Lange, Kerstin Strupat, and Stevan Horning. Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Anal Chem*, 78(7):2113–20, Apr 2006.

[90] Qizhi Hu, Robert J. Noll, Hongyan Li, Alexander Makarov, Mark Hardman, and R. Graham Cooks. The orbitrap: a new mass spectrometer. *Journal of Mass Spectrometry*, 40(4):430–443, 2005.

[91] Amelia C. Peterson, Jason D. Russell, Derek J. Bailey, Michael S. Westphall, and Joshua J. Coon. Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Molecular Cellular Proteomics*, 11(11):1475–1488, Aug 2012.

[92] Vagisha Sharma, Josh Eckels, Birgit Schilling, Christina Ludwig, Jacob D Jaffe, Michael J MacCoss, and Brendan MacLean. Panorama public: A public repository for quantitative data sets processed in skyline. *Mol Cell Proteomics*, 17(6):1239–1244, Jun 2018.

[93] Shao-En Ong and Matthias Mann. Mass spectrometry–based proteomics turns quantitative. *Nature Chemical Biology*, 1(5):252–262, Oct 2005.

[94] Barbara Deracinois, Christophe Flahaut, Sophie Duban-Deweer, and Yannis Karamanos. Comparative and quantitative global proteomics approaches: An overview. *Proteomes*, 1(3):180–218, Oct 2013.

[95] M. Wang, J. You, K. G. Bemis, T. J. Tegeler, and D. P. G. Brown. Label-free mass spectrometry-based protein quantification technologies in proteomic analysis. *Briefings in Functional Genomics and Proteomics*, 7(5):329–339, Jun 2008.

[96] Marcus Bantscheff, Markus Schirle, Gavain Sweetman, Jens Rick, and Bernhard Kuster. Quantitative mass spectrometry in proteomics: a critical review. *Analytical and Bioanalytical Chemistry*, 389(4):1017–1031, Aug 2007.

[97] Shao-En Ong, Blagoy Blagoev, Irina Kratchmarova, Dan Bach Kristensen, Hanno Steen, Akhilesh Pandey, and Matthias Mann. Stable isotope labeling by amino acids in cell culture, silac, as a simple and accurate approach to expression proteomics. *Molecular Cellular Proteomics*, 1(5):376–386, May 2002.

[98] Steven P. Gygi, Beate Rist, Scott A. Gerber, Frantisek Turecek, Michael H. Gelb, and Ruedi Aebersold. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnology*, 17:994 EP –, 10 1999.

[99] Philip L. Ross, Yulin N. Huang, Jason N. Marchese, Brian Williamson, Kenneth Parker, Stephen Hattan, Nikita Khainovski, Sasi Pillai, Subhakar Dey, Scott Daniels, and et al. Multiplexed protein quantitation insaccharomyces cerevisiaeusing amine-reactive isobaric tagging reagents. *Molecular Cellular Proteomics*, 3(12):1154–1169, Sep 2004.

[100] Andrew Thompson, Jürgen Schäfer, Karsten Kuhn, Stefan Kienle, Josef Schwarz, Günter Schmidt, Thomas Neumann, R Johnstone, A Karim A Mohammed, and Christian Hamon. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by ms/ms. *Anal Chem*, 75(8):1895–904, Apr 2003.

[101] Daniel Kolarich, Pia H Jensen, Friedrich Altmann, and Nicolle H Packer. Determination of site-specific glycan heterogeneity on glycoproteins. *Nature Protocols*, 7(7):1285–1298, Jun 2012.

[102] William R Alley, Jr, Benjamin F Mann, and Milos V Novotny. High-sensitivity analytical approaches for the structural characterization of glycoproteins. *Chem Rev*, 113(4):2668–732, Apr 2013.

[103] Yvonne Pasing, Albert Sickmann, and Urs Lewandrowski. N-glycoproteomics: mass spectrometry-based glycosylation site annotation. *Biological Chemistry*, 393(4), Jan 2012.

[104] Manfred Wuhrer, M. Isabel Catalina, André M. Deelder, and Cornelis H. Hokke. Glycoproteomics based on tandem mass spectrometry of glycopeptides. *Journal of Chromatography B*, 849(1-2):115–128, Apr 2007.

[105] Qing Yu, Bowen Wang, Zhengwei Chen, Go Urabe, Matthew S. Glover, Xudong Shi, Lian-Wang Guo, K. Craig Kent, and Lingjun Li. Electron-transfer/higher-energy collision dissociation (ethcd)-enabled intact glycopeptide/glycoproteome characterization. *Journal of The American Society for Mass Spectrometry*, 28(9):1751–1764, Jul 2017.

[106] Shiaw-Lin Wu, Andreas F R Hühmer, Zhiqi Hao, and Barry L Karger. On-line lc-ms approach combining collision-induced dissociation (cid), electron-transfer dissociation (etd), and cid of an isolated charge-reduced species for the trace-level characterization of proteins with post-translational modifications. *J Proteome Res*, 6(11):4230–44, Nov 2007.

[107] Peng Zhao, Rosa Viner, Chin Fen Teo, Geert-Jan Boons, David Horn, and Lance Wells. Combining high-energy c-trap dissociation and electron transfer dissociation for protein o-glcnac modification site assignment. *J Proteome Res*, 10(9):4088–104, Sep 2011.

[108] Jolene K. Diedrich, Antonio F. M. Pinto, and John R. Yates. Energy dependence of hcd on peptide fragmentation: Stepped collisional energy finds the sweet spot. *Journal of The American Society for Mass Spectrometry*, 24(11):1690–1699, Aug 2013.

[109] Hong Yang, Chenxi Yang, and Taolei Sun. Characterization of glycopeptides using a stepped higher-energy c-trap dissociation approach on a hybrid quadrupole orbitrap. *Rapid Communications in Mass Spectrometry*, 32(16):1353–1362, Jul 2018.

[110] Charandeep Singh, Cleidiane G Zampronio, Andrew J Creese, and Helen J Cooper. Higher energy collision dissociation (hcd) product ion-triggered electron transfer dissociation (etd) mass spectrometry for the analysis of n-linked glycoproteins. *J Proteome Res*, 11(9):4517–25, Sep 2012.

[111] Christian K Frese, A F Maarten Altelaar, Henk van den Toorn, Dirk Nolting, Jens Griep-Raming, Albert J R Heck, and Shabaz Mohammed. Toward full peptide sequence coverage by dual fragmentation combining electron-transfer and higher-energy collision dissociation tandem mass spectrometry. *Anal Chem*, 84(22):9668–73, Nov 2012.

[112] Benjamin L. Parker, Morten Thaysen-Andersen, Daniel J. Fazakerley, Mira Holliday, Nicolle H. Packer, and David E. James. Terminal galactosylation and sialylation switching on membrane glycoproteins upon tnf-alpha-induced insulin resistance in adipocytes. *Molecular Cellular Proteomics*, 15(1):141–153, Nov 2015.

[113] Marshall Bern, Yong J Kil, and Christopher Becker. Byonic: advanced peptide and protein identification software. *Curr Protoc Bioinformatics*, Chapter 13:Unit13.20, Dec 2012.

[114] Li Cao, Nikola Tolić, Yi Qu, Da Meng, Rui Zhao, Qibin Zhang, Ronald J Moore, Erika M Zink, Mary S Lipton, Ljiljana Paša-Tolić, and Si Wu. Characterization of intact n- and o-linked glycopeptides using higher energy collisional dissociation. *Anal Biochem*, 452:96–102, May 2014.

[115] Johannes Stadlmann, Jasmin Taubenschmid, Daniel Wenzel, Anna Gattinger, Gerhard Dürnberger, Frederico Dusberger, Ulrich Elling, Lukas Mach, Karl Mechtler, and Josef M. Penninger. Comparative glycoproteomics of stem cells identifies new players in ricin toxicity. *Nature*, 549(7673):538–542, Sep 2017.

[116] J Hall, G P Hazlewood, M A Surani, B H Hirst, and H J Gilbert. Eukaryotic and prokaryotic signal peptides direct secretion of a bacterial endoglucanase by mammalian cells. *Journal of Biological Chemistry*, 265(32):19996–9, 1990.

[117] Y. C. Tsai and A. M. Weissman. The unfolded protein response, degradation from the endoplasmic reticulum, and cancer. *Genes Cancer*, 1(7):764–778, Jul 2010.

[118] Shengzhan Luo, Peter Baumeister, Shujie Yang, Steve F. Abcouwer, and Amy S. Lee. Induction of grp78/bip by translational block. *Journal of Biological Chemistry*, 278(39):37375–37385, Jul 2003.

[119] Natalia A. Cherepanova and Reid Gilmore. Mammalian cells lacking either the cotranslational or posttranslocational oligosaccharyltransferase complex display substrate-dependent defects in asparagine linked glycosylation. *Scientific Reports*, 6(1), Feb 2016.

[120] Edward Avezov, Efrat Ron, Yana Izenshtein, Yosef Adan, and Gerardo Z. Lederkremer. Pulse-chase analysis of n-linked sugar chains from glycoproteins in mammalian cells. *Journal of Visualized Experiments*, (38), Apr 2010.

[121] Haiping Wang, Xin Wang, Zun-Ji Ke, Ashley L Comer, Mei Xu, Jacqueline A Frank, Zhuo Zhang, Xianglin Shi, and Jia Luo. Tunicamycin-induced unfolded protein response in the developing mouse brain. *Toxicol Appl Pharmacol*, 283(3):157–67, Mar 2015.

[122] J Alan Diehl, Serge Y Fuchs, and Costantinos Koumenis. The cell biology of the unfolded protein response. *Gastroenterology*, 141(1):38–41.e412, 07 2011.

[123] Miao Wang and Randal J. Kaufman. The impact of the endoplasmic reticulum protein-folding environment on cancer development. *Nature Reviews Cancer*, 14(9):581–597, Sep 2014.

[124] Afshin Samali, Una FitzGerald, Shane Deegan, and Sanjeev Gupta. Methods for monitoring endoplasmic reticulum stress and the unfolded protein response. *International Journal of Cell Biology*, 2010:1–11, 2010.

5 Supplementary data

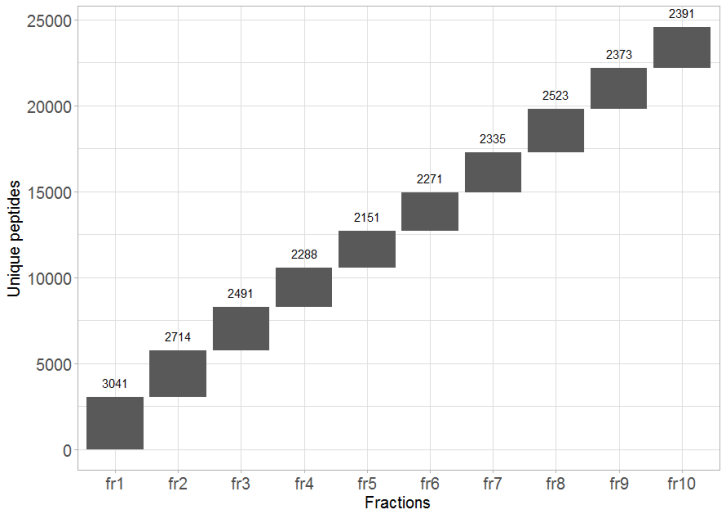


Figure S1: STT3A KO - Distribution of identified unique peptides in 10 fractions of a high pH chromatographic separation

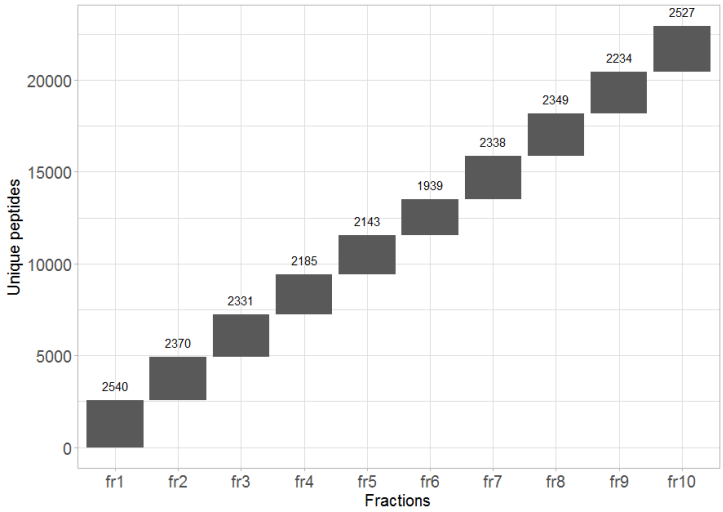


Figure S2: STT3B KO - Distribution of identified unique peptides in 10 fractions of a high pH chromatographic separation

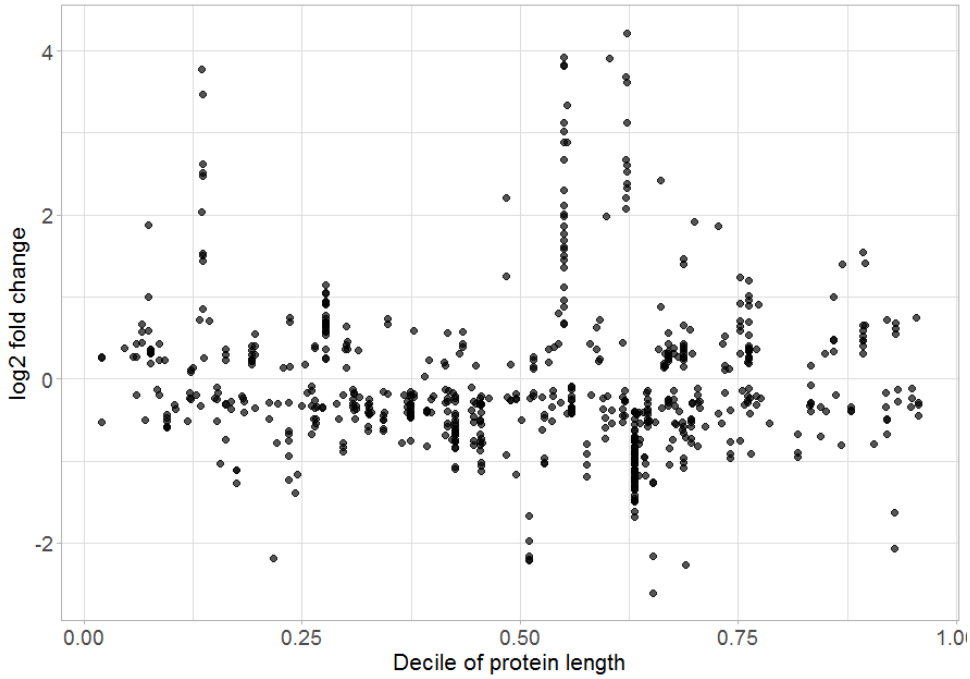


Figure S3: STT3A KO - Glycosylation distribution over the whole dataset

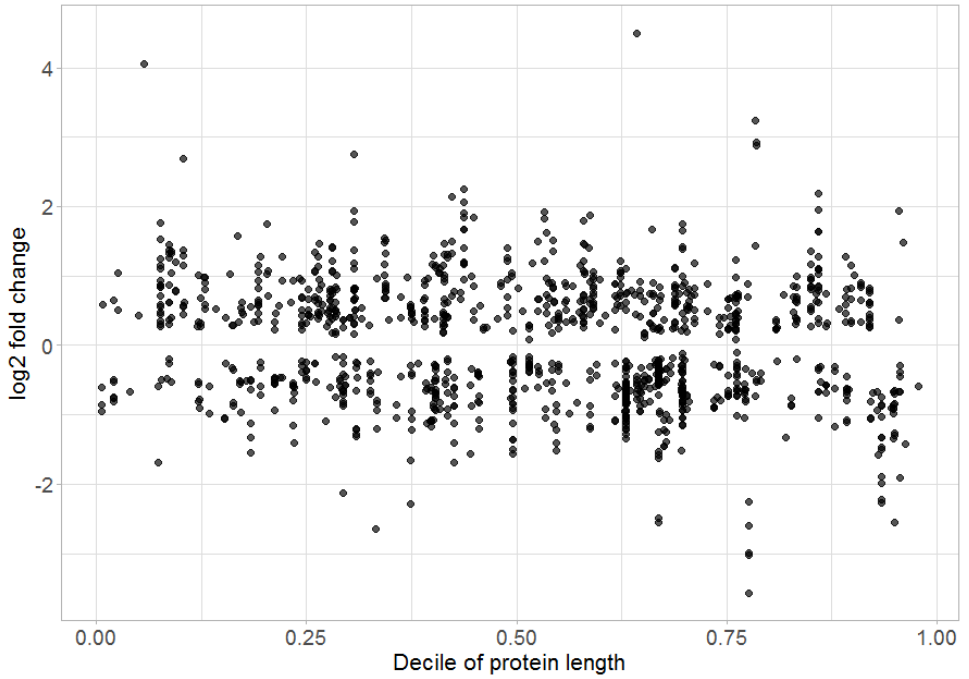


Figure S4: STT3B KO - Glycosylation distribution over the whole dataset

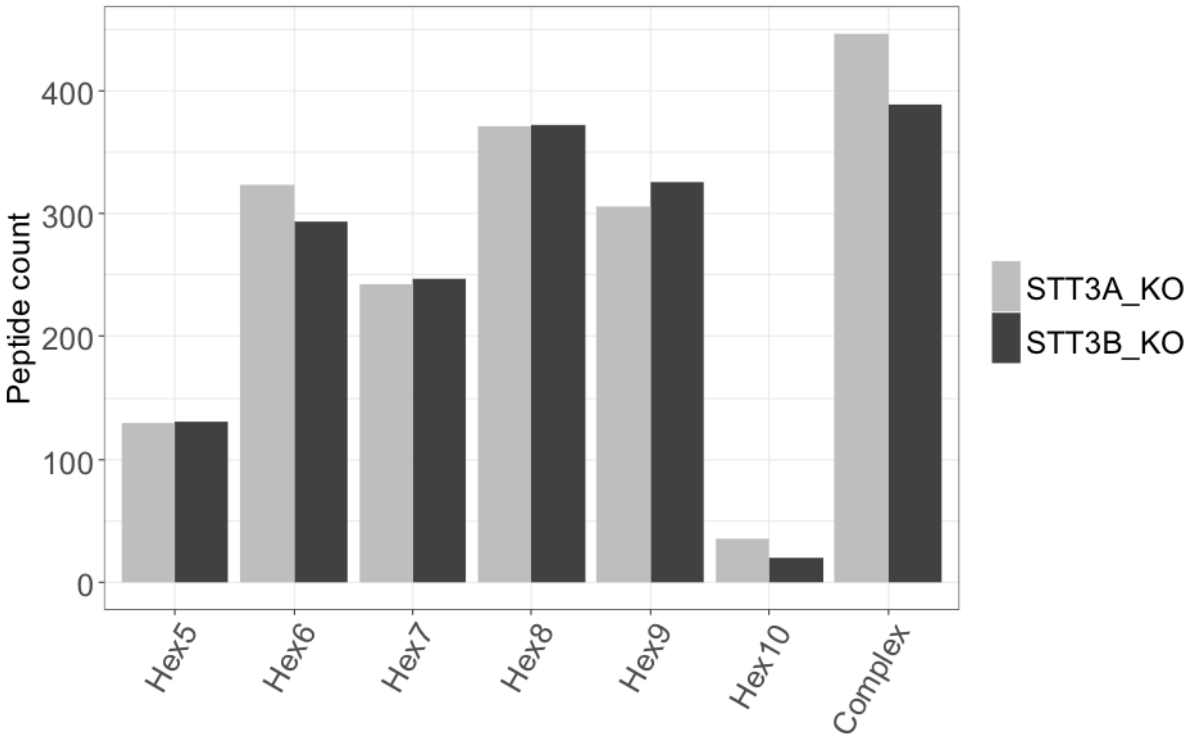


Figure S5: Number of identified peptide comprising different glycan structures in both KO states

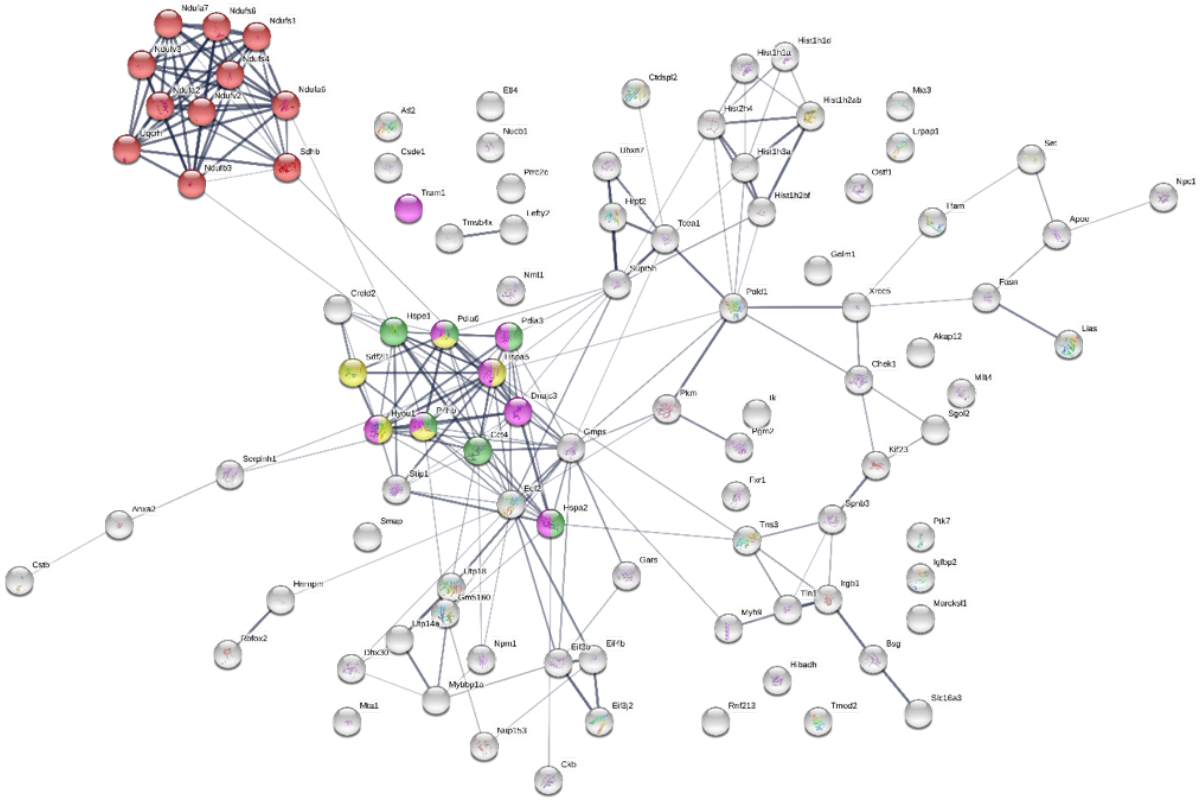


Figure S6: STT3A KO - upregulated proteins at p=0.05 and foldchange > 1; in green, yellow and purple = ER processing and stress response; in red = electron transport chain

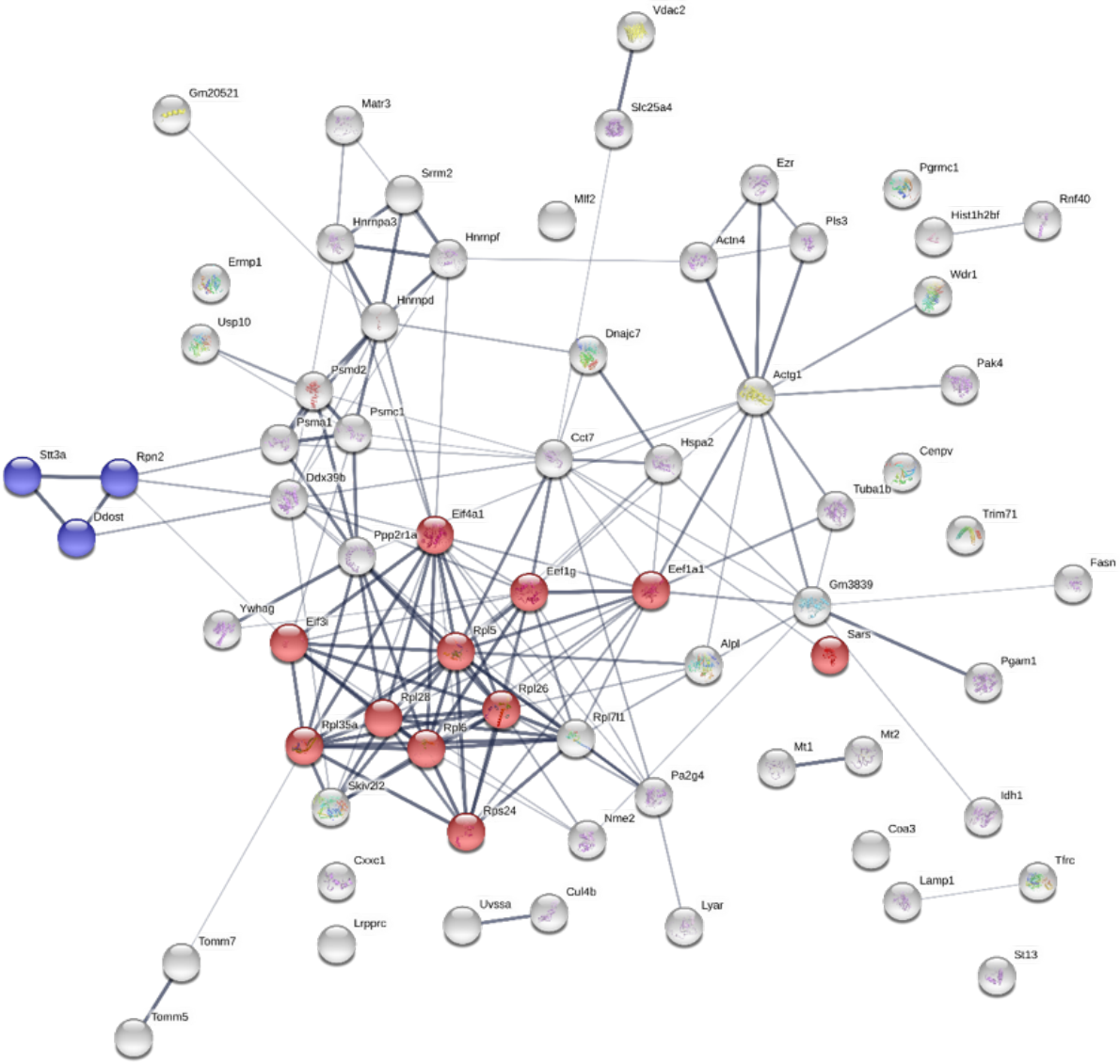


Figure S7: STT3A KO - downregulated proteins at p=0.05 and foldchange > 1; in red= translation; in blue= N-glycan biosynthesis

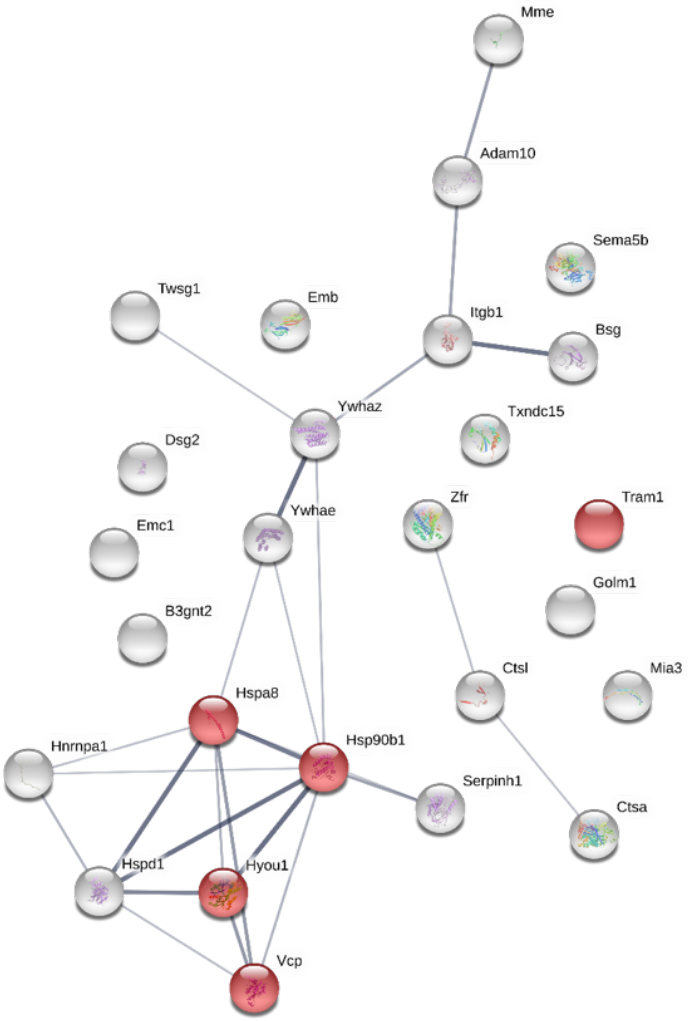


Figure S8: STT3A KO - upregulated glycoproteins at p=0.05 and foldchange $>$ 1; in red= ER protein processing

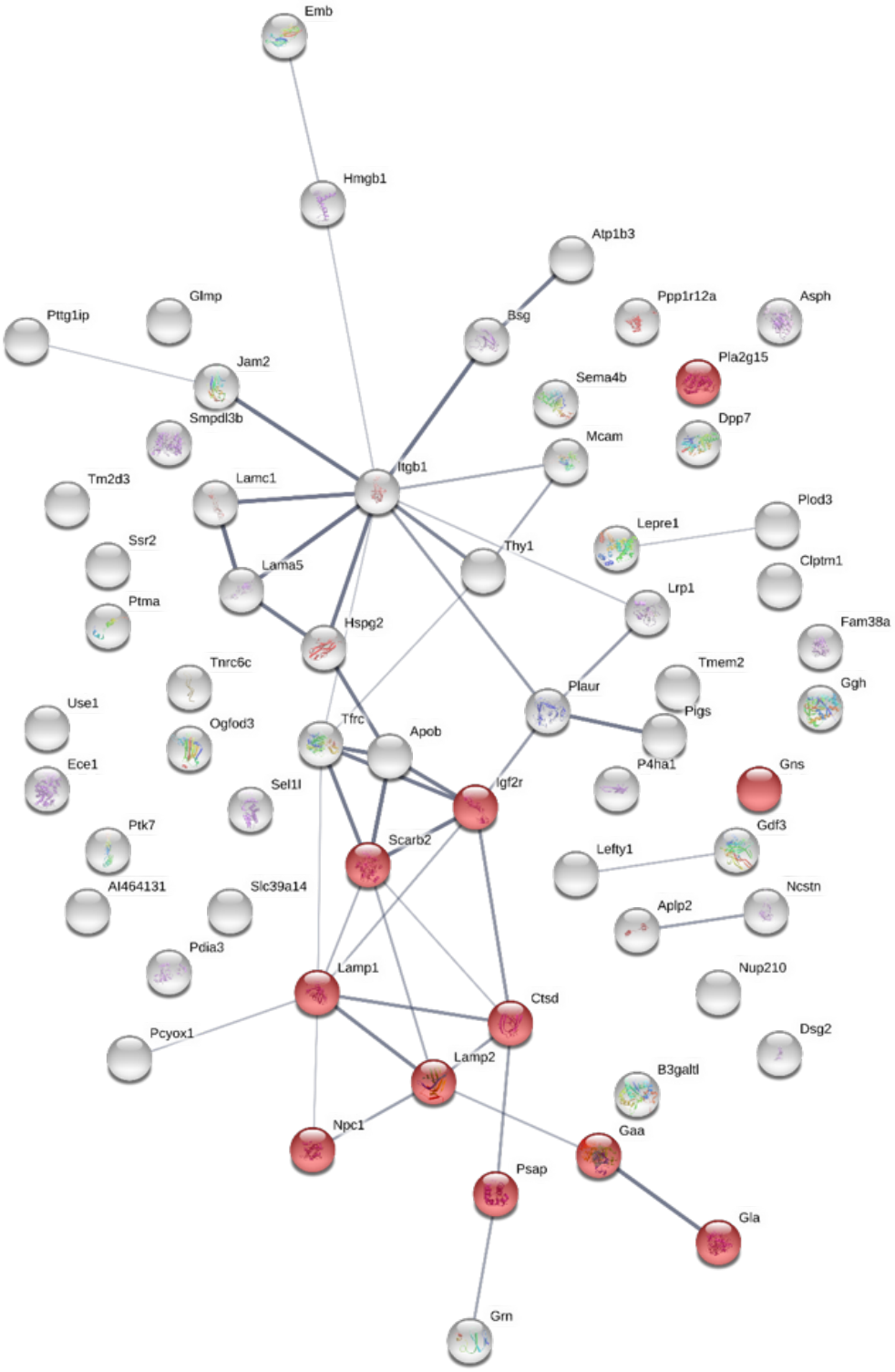


Figure S9: STT3A KO - downregulated glycoproteins at p=0.05 and foldchange > 1; red= lysosomal proteins

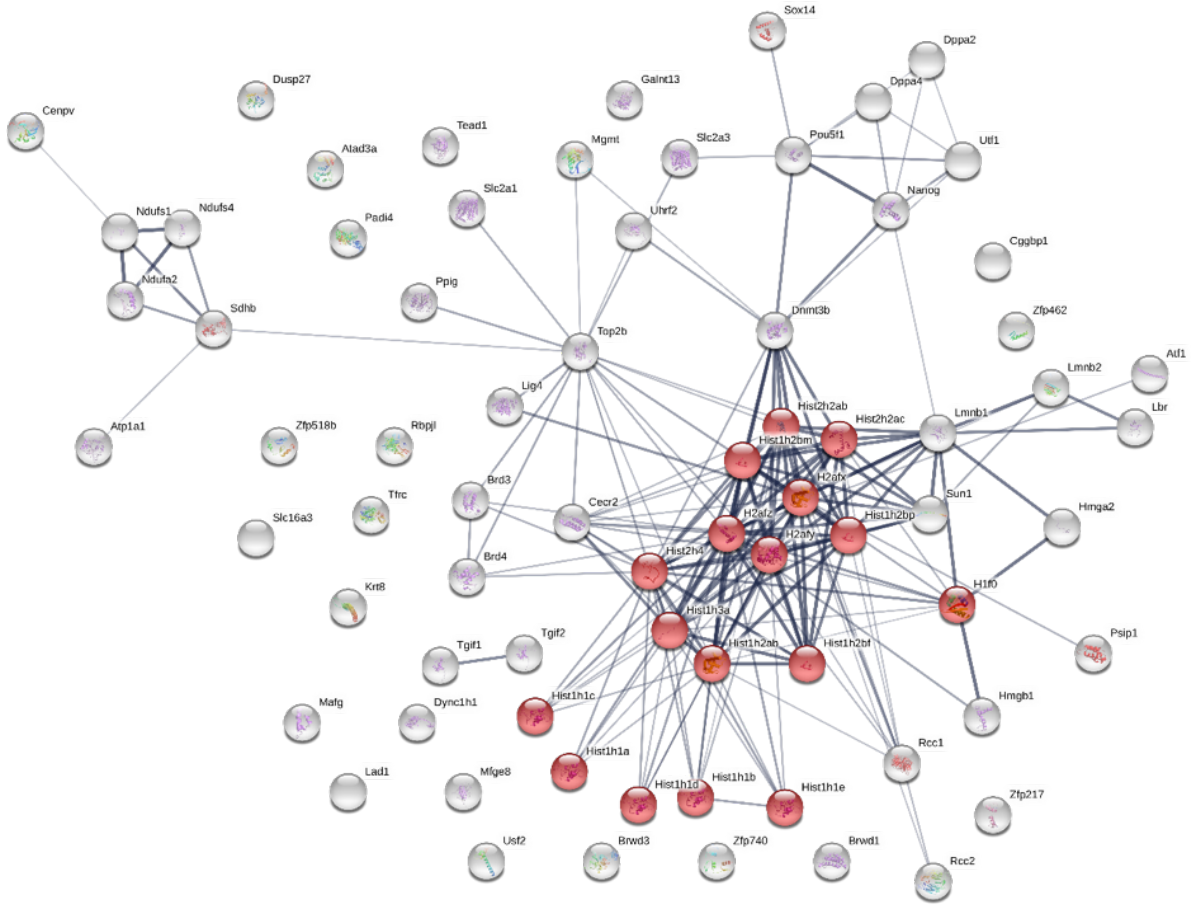


Figure S10: STT3B KO - upregulated proteins at p=0.05 and foldchange > 1; in red= nucleosomal proteins

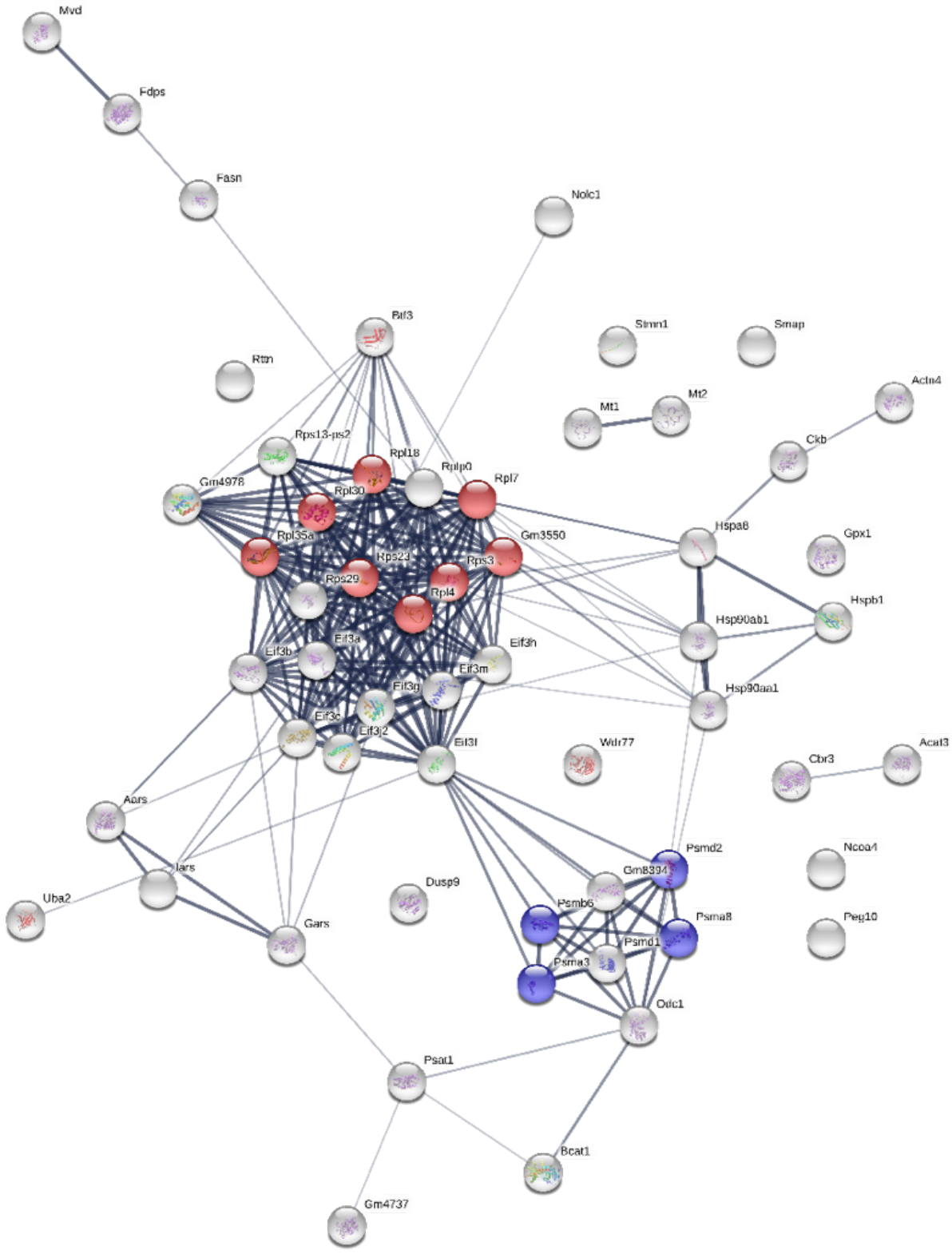


Figure S11: STT3B KO - downregulated proteins at p=0.05 and foldchange > 1; in red= ribosomal proteins; in blue= proteasome proteins

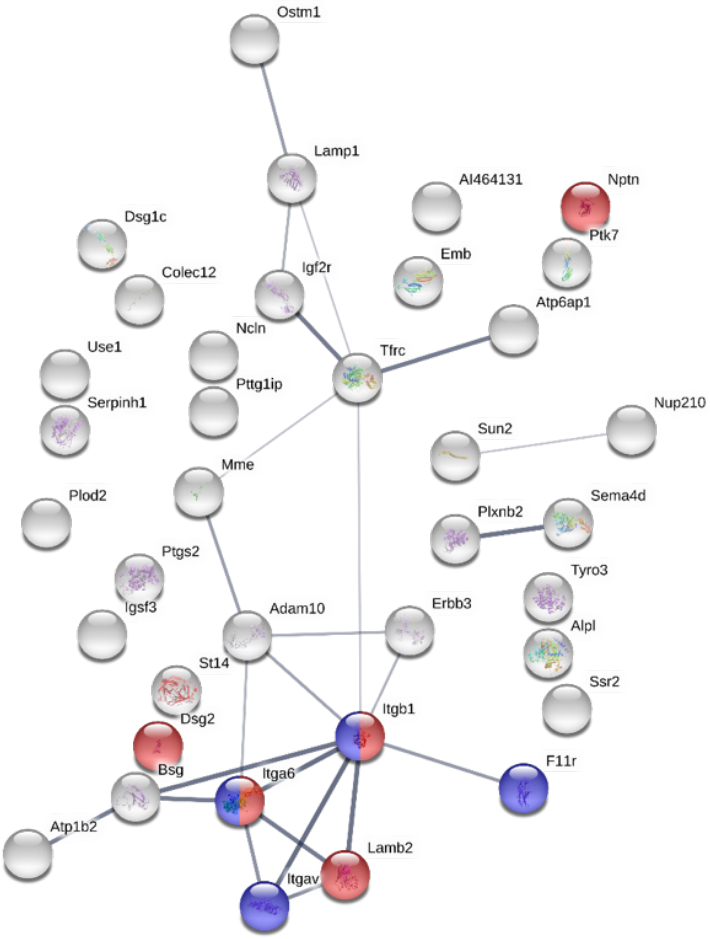


Figure S12: STT3B KO - upregulated glycoproteins at p=0.05 and foldchange > 1; in red and blue= CAMs

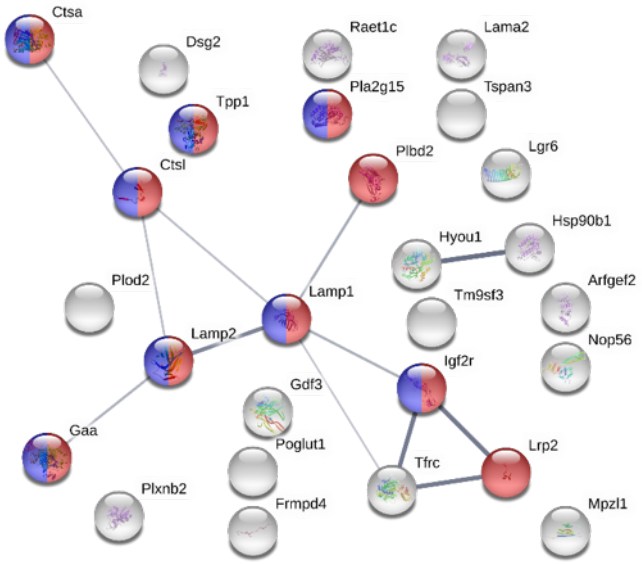


Figure S13: STT3B KO - downregulated glycoproteins at p=0.05 and foldchange > 1; in red and blue= lysosomal proteins

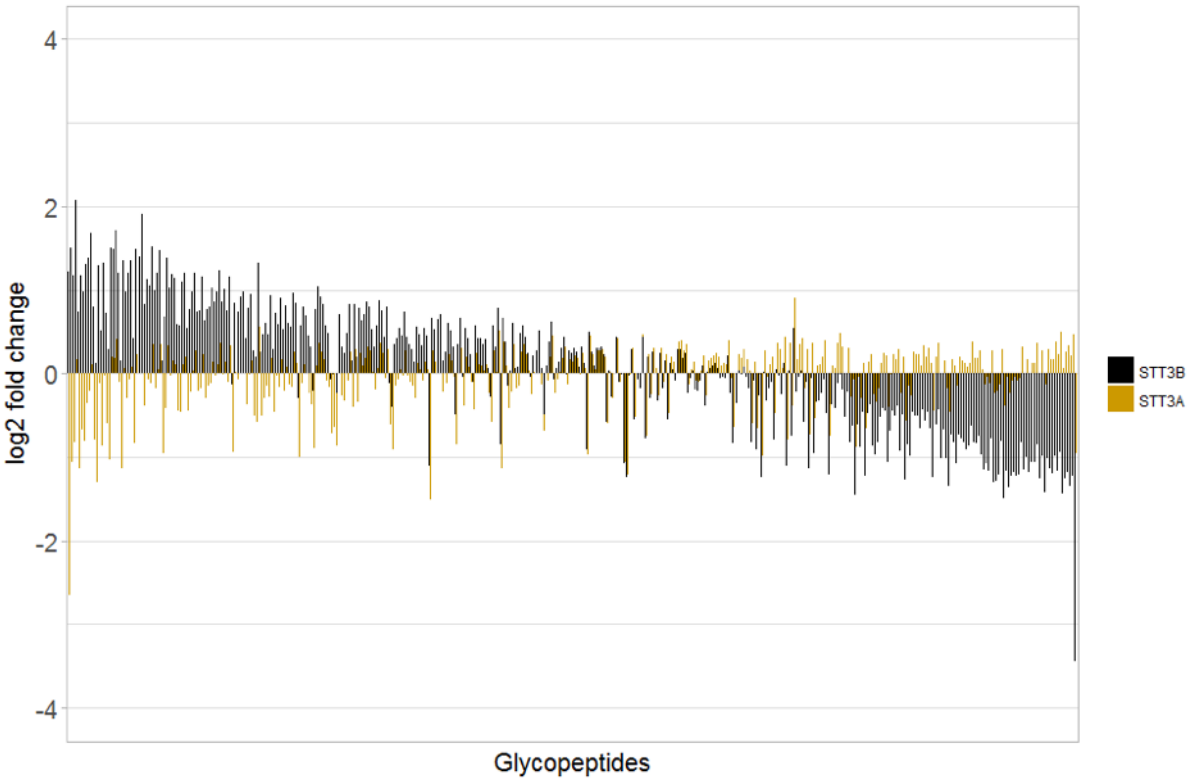


Figure S14: Sequon analysis on glycoproteomics data set

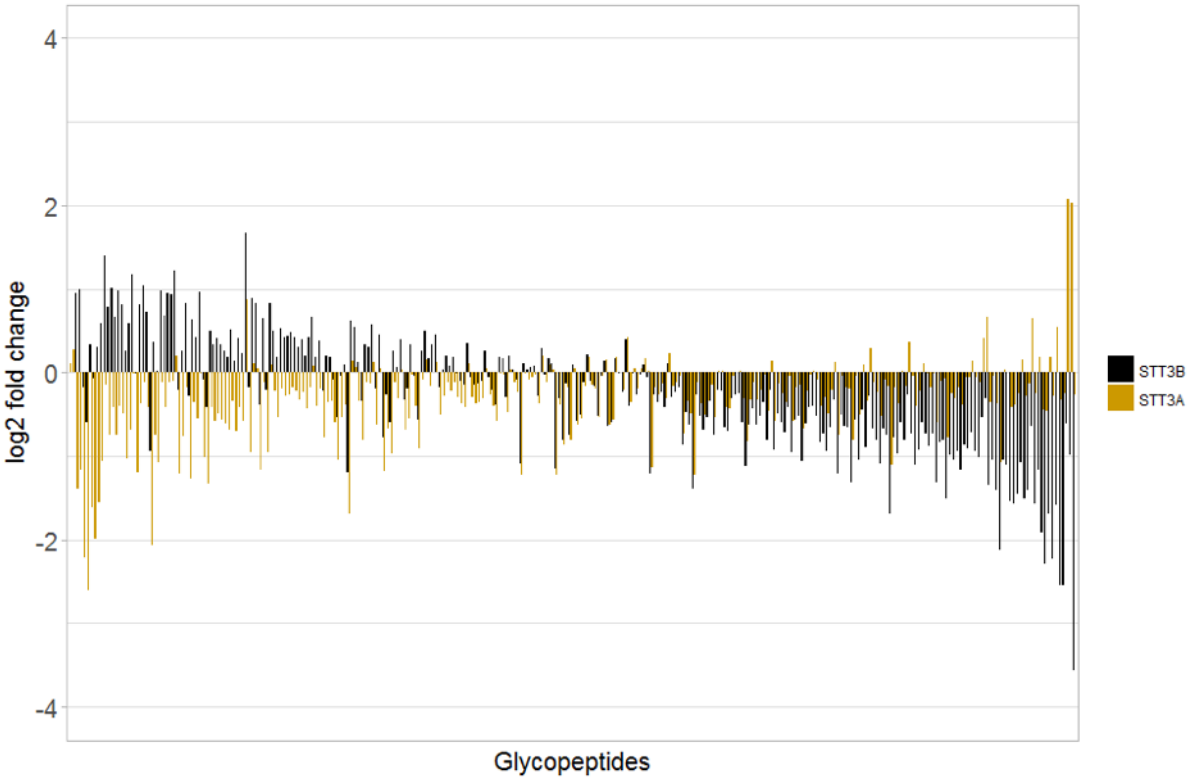


Figure S15: Sequon analysis on proteomics data set