

Systems-wide analysis of glycoprotein conformational changes by limited deglycosylation assay

Simon Ngao Mule¹, Livia Rosa-Fernandes¹, João V. P. Coutinho¹, Vinícius De Morais^{1,2}, Janaina Macedo da Silva¹, Verônica Feijoli Santiago¹, Daniel Quina¹, Gilberto Santos de Oliveira¹, Morten Thaysen-Andersen³, Martin R. Larsen⁴, Letícia Labriola², Giuseppe Palmisano^{1#}

1. GlycoProteomics Laboratory, Department of Parasitology, Institute of Biomedical Sciences, University of Sao Paulo, Sao Paulo, Brazil
2. Department of Biochemistry, Institute of Chemistry, University of Sao Paulo, Sao Paulo, Brazil
3. Department of Molecular Sciences, Macquarie University, Sydney, Australia
4. Department of Biochemistry and Molecular Biology, University of Southern Denmark, Odense, DK

To whom correspondence should be addressed: Prof. Giuseppe Palmisano, Glycoproteomics Laboratory Department of Parasitology, ICB, University of São Paulo, Brazil, Av. Prof. Lineu Prestes, 1374, 05508-900 - São Paulo – SP – Brazil, Tel: + 55-11-3091-3899, palmisano.gp@gmail.com, palmisano.gp@usp.br

Abstract

A new method to probe the conformational changes of glycoproteins on a systems-wide scale, termed limited deglycosylation assay (LDA), is described. The method measures the differential rate of deglycosylation of N-glycans on natively folded proteins by the common peptide:N-glycosidase F (PNGase F) enzyme which in turn informs on their spatial presentation and solvent exposure on the protein surface hence ultimately the glycoprotein conformation. LDA involves 1) protein-level N-deglycosylation under native conditions, 2) trypsin digestion under denaturing conditions, 3) glycopeptide enrichment, 4) peptide-level N-deglycosylation and 5) quantitative MS-based analysis of the formerly N-glycosylated peptides. LDA was initially developed and the experimental conditions optimized using bovine RNase B and fetuin. The method was then applied to glycoprotein extracts from LLC-MK2 epithelial cells upon treatment with dithiothreitol to induce endoplasmic reticulum stress and promote protein misfolding. Data from the LDA and 3D structure analysis showed that glycoproteins predominantly undergo structural changes in loops/turns upon ER stress as exemplified with detailed analysis of ephrin-A5, GALNT10, PVR and BCAM. These results show that LDA accurately reports on systems-wide conformational changes of glycoproteins induced under controlled treatment regimes. Thus, LDA opens avenues to study glycoprotein structural changes in a range of other physiological and pathophysiological conditions relevant to acute and chronic diseases.

Keywords

Glycoprotein; PNGase F; glycoproteomics; mass spectrometry; ER stress; protein folding; N-glycosylation; deglycosylation.

Introduction

In the endoplasmic reticulum (ER), proteins are folded to their three-dimensional native conformational states depending on their amino acid sequence [1]. Proteins are folded by a complex molecular network composed of chaperones and protein folding enzymes regulated under physiological and stress conditions [2-4]. However, intracellular stress can impair ER quality control (ERQC) of proteins destined for secretion or transport to membranes, leading to the accumulation of misfolded and/or unfolded inactive precursor proteins in the ER causing ER stress [5-7]. Several diseases have been associated with a dysregulation of protein folding and accumulation of misfolded aggregates, including amyloidosis (e.g. Alzheimer's disease) and prion diseases [8-10]. Protein kinase PKR-like ER kinase, inositol-requiring protein 1 alpha and activating transcription factor 6 alpha sense dysregulation of protein folding in the ER, leading to the activation of a network of signalling pathways termed the unfolded protein response (UPR), an evolutionary conserved mechanism serving to restore ER homeostasis and refold or degrade the unfolded proteins [11, 12]. UPR includes the downregulation of protein translation [13, 14], upregulation of protein-folding enzymes and chaperones [15], and upregulation of components of the ER-associated degradation (ERAD) pathway to enhance the degradation of the misfolded proteins [16, 17]. Proteins which cannot be refolded are targeted to the ERAD pathway, where they are ubiquitinated and degraded via the proteasomal pathway [18, 19]. In conditions of excess or sustained ER stress, autophagy can also be induced. If this process cannot revert the damage, it will end up triggering apoptosis culminating in cell death [20]. Chemicals such as tunicamycin (TM), dithiothreitol (DTT) and thapsigargin, which trigger ER stress through distinct mechanisms, have been used to study UPR [21, 22].

Newly synthesized proteins are co- and post-translationally modified with glycans in the ER followed by subsequent processing in the Golgi. This post-translational modification (PTM) is important for the correct folding and functioning of the proteins. Glycosylation is a vital and ubiquitous PTM, which, in eukaryotes, modifies over half of the proteome [23]. The *en bloc* transfer of pre-assembled oligosaccharides from a lipid-linked oligosaccharide donor to proteins in the ER at asparagine

residues (Asn/N) in the conserved sequon (Asn(Xxx)Ser/Thr/Cys; where Xxx \neq Pro) and the subsequent processing and maturation in the Golgi is the first step of the N-glycosylation pathway [24]. Protein glycosylation and the repertoire of glycans carried by newly synthesized glycoproteins is determined by the nature of the proteins undergoing glycosylation [25, 26] as well as a host of factors of the complex glycosylation machinery including the level of glycosylation enzymes [27], the availability of sugar nucleotide donors, subcellular localization [28], and cellular metabolism [29]. Glycans play diverse and important biological and structural functions including impacting the protein folding and localization, cell adhesion, function and regulation, stability of secreted proteins, protein solubility, glycan dependent regulation of receptor signalling and host-pathogen interaction [30-38]. The dysregulation of protein glycosylation has been associated with several human diseases including congenital disorders of glycosylation, diabetes mellitus, infectious diseases, auto-immune diseases, prion disease, cancer, cardiac rhythm disorders and chronic inflammatory diseases [39-41]. N-glycan dependent quality control of glycoproteins is regulated by lectin chaperones such as calnexin and calreticulin, disulphide isomerase, glucosyltransferases and glucosidases that prevent unfolded, misfolded and aggregated glycoproteins to exit the ER and enter the Golgi compartments. The presence of specific N-glycan structures within the ER correlate with the folding status and age of glycoproteins directing the activity of the folding machinery [42, 43].

X-ray crystallography and two-dimensional nuclear magnetic resonance (NMR) spectroscopy techniques are the major techniques applied to elucidate macromolecule structures including proteins [44, 45], and more recently, cryo-electron microscopy (cryoEM) [46]. These techniques, however, suffer technological setbacks including the need of highly purified protein crystals for X-ray crystallography, the need for high concentrations of non-aggregating proteins for NMR [47], and the need for sample structural homogeneity for cryoEM. The application of these techniques for glycoproteins is hindered by the size and heterogeneity of the oligosaccharide chains and the surface entropy of this PTM which impacts the crystallization. Limited proteolysis (LiP) is an alternative method for protein structure analysis of complex biological samples at scale [48]. LiP probes the conformation of proteins by employing peptidases which perform limited proteolysis on surface exposed peptide bonds of natively folded proteins [49]. The most recent development of LiP-MS is based on a dual-digestion using firstly a non-specific protease applied under native conditions followed by trypsin digestion under denaturing conditions. The resulting peptides are analysed by quantitative shotgun or targeted mass spectrometry enabling the identification of the

structure-dependent proteolytic patterns [49]. This method has not explored the surface accessibility of glycoproteins.

We have here developed a new method for the large-scale analysis of glycoprotein conformational changes using limited deglycosylation assay (LDA). Previously it was shown that cells overexpressing endogenous PNGase F could be used as a sensor for studying the folding state of specific glycoproteins measured by differential SDS-PAGE migration [50, 51]. In the LDA method described herein, we have extended the concepts of limited proteolysis and deglycosylation sensitivity of differentially folded glycoproteins using large-scale differential and limited PNGase F deglycosylation.

Materials and Methods

Proof-of-concept and method optimization

Proof-of-concept and optimization of the LDA method were performed using two standard glycoproteins, bovine RNase B and fetuin (New England Biolabs) by PNGase F (New England Biolabs) under native conditions. The subsequent detection of the partially deglycosylated glycoproteins were performed by differential band migration on 15% SDS-PAGE. Detergent concentration (Nonidet P-40/NP-40) in the lysis buffer (adopted from [51]), PNGase F to glycoprotein ratio, incubation temperature and time were optimized. Final concentrations of 0.4% and 1% (v/v) NP-40 in the lysis buffer were tested. The optimal PNGase F (Cat. P0709S, NEB) concentration (relative to the glycoprotein substrate) was tested ranging from 1000, 500, 250, 125 to 75 U/ 5 µg RNase B (Cat.P7817S, NEB) and fetuin (Cat. P6042S, NEB). Different incubation times (2, 4, 8 and 16 h) were assayed, in addition to three temperature points (4°C, 24°C and 37°C). In all optimization conditions, fully deglycosylated and undeglycosylated RNase B and fetuin were included as controls. The fully denatured standard glycoproteins were prepared as recommended by the manufacturer (NEBs). Briefly, 5 µg standard glycoprotein was mixed in 1 µL of 10× glycoprotein denaturing buffer (NEB) (final conc: 0.5% SDS, 40 mM DTT) and milli-Q ultrapure[®] H₂O to 10 µL final reaction volume and boiled for 10 min at 95°C, followed by sample cooling on ice. Subsequently, deglycosylation of the denatured glycoproteins was performed in a 20 µL reaction volume consisting of 10 µL of the previously denatured glycoprotein, 2 µL 10% NP-40 (final concentration 1% NP-40), 2 µL 10X GlycoBuffer 2 (final concentration 50 mM sodium phosphate (pH 7.5), 250 U PNGase F and 6 µL milli-Q ultrapure H₂O. Deglycosylation was performed at 37°C for 1 h. Deglycosylation on the

standard proteins was analyzed by gel electrophoresis on 15% SDS-PAGE gels and the gels stained by Coomassie Brilliant Blue R-250 Dye (Cat. 20278, Thermo Scientific). The optimal conditions were applied to study limited deglycosylation on LLC-MK2 cellular glycoproteins upon induction of ER stress, as described below.

Cell cultures

Rhesus monkey kidney epithelial cells (LLC-MK2) (ATCC, USA) [52] were maintained in RPMI media (Life Technologies) supplemented with 10% (v/v) heat-inactivated fetal bovine serum (FBS) and 100 U/mL penicillin in 75 cm² culture flasks at 37°C with 5% CO₂ until the cells reached 90% confluence. Passaging was done by seeding 1.0×10^6 cells in new 75 cm² culture flasks. Cell cultures were periodically assayed for mycoplasma contamination using PCR test.

ER stress induction, cell lysis under native conditions and protein extraction

LLC-MK2 cells were seeded and grown in RPMI media supplemented with 10% FBS at 37°C with 5% CO₂ to 90% confluence. The cells were subsequently rinsed 3 times with 1× PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄, pH 7.4), treated with 3 mM dithiothreitol (DTT) (molecular grade, Cat. 3483-12-3, Promega), and incubated for 30 min at 37°C in 5% CO₂ in serum-free RPMI media. This concentration of DTT has previously been shown to induce ER stress in several cell lines [53-55]. Control conditions of LLC-MK2 cells incubated in the absence of DTT were included. After incubation, the cells were washed 3 times with ice-cold 1 × PBS.

Subsequently, the cells were detached by cell scraping technique in 2 mL ice-cold 1× PBS, and transferred into pre-chilled Protein LoBind® 2 mL microtubes (Eppendorf AG, Hamburg, Germany). The cells were pelleted by low-speed centrifugation at 800×g for 3 min at 4°C. The supernatant was discarded by aspiration, and the cells were resuspended in 200 µL ice-cold cell lysis buffer adopted from [50, 51] with minor modifications (50 mM Tris/HCl (pH 8.0), 0.4% NP-40, and 150 mM NaCl) containing protease inhibitor cocktail (1×) (Promega), followed by cell lysis by incubation on ice for 20 min. Ultimately, the lysed cell debris were cleared by centrifugation at 17,800×g for 10 min at 4°C and the supernatants transferred separately into new pre-chilled Protein LoBind® microtubes placed on ice. The samples were submitted to total proteome analysis and LDA. Proteins were assayed using SDS-PAGE, and the proteome from both control and ER stress-induced fractions submitted to sample processing prior to label free nLC-MS/MS analysis as described below.

To determine the protein concentration, cells were cultured as previously described and resuspended in 8 M urea supplemented with 1X protease inhibitors, lysed by 3 freeze thaw cycles, cleared by centrifugation as before, and the proteins quantified by Qubit fluorometric detection method.

Optimized LDA on cellular proteins

The optimized conditions previously determined were applied to LLC-MK2 cellular glycoproteins and confirmed by lectin blot analysis prior to nLC-MS/MS analysis. These conditions were: PNGase F (1000 U/100 µg protein), 16 h incubation at 4°C in 200 µL lysis buffer with 0.4% NP-40. Biological replicates for each condition with and without PNGase F treatment were analyzed. PNGase F was deactivated by boiling the samples for 10 min at 95°C in Laemmli sample buffer (final concentration of 31.5 mM Tris, pH 6.8, 10% glycerol, 1% SDS, 10 mM DTT, 0.005% bromophenol blue) [56].

SDS-PAGE and lectin blotting

To visualize the effect of DTT on (glyco)protein expression, 10 µg protein extract from LLC-MK2 cells treated with or without 3 mM DTT was separated in 12% SDS-PAGE at a constant voltage of 125 V, stained by CBB, destained and the protein profiles acquired using a ChemiDoc Imaging system (Bio-Rad). Limited deglycosylation of LLC-MK2 cellular glycoproteins treated with or without 3 mM DTT, and treated with or without PNGase F was assayed by lectin blotting using Concanavalin A (Con A) (Cat. B-1005-5, Vector Laboratories). A total of 10 µg protein each were electrophoresed in a 12% SDS-PAGE gel and transferred onto PVDF membranes at a constant of 300 mA for 2 h. The membranes were blocked with 5% BSA and probed overnight at 4°C with biotinylated Con A lectin followed by horseradish peroxidase (HRP) secondary antibody for 1 h at RT with orbital shaking. The lectin blots were developed using SuperSignal™ West Pico Plus Chemiluminescent substrate (Thermo Fischer Scientific) and the images acquired using a ChemiDoc Imaging system (Bio-Rad). The band intensities were analyzed using the ImageLab (BioRad) software. Statistical analyses of the lectin intensities were calculated in GraphPad (v.8) using one-way ANOVA followed by Tukey's post-hoc test, with a $p < 0.05$ considered statistically significant.

Sample processing, protein digestion and peptide extraction

To facilitate removal of NP-40 detergent prior to LC-MS/MS analysis, we performed in-gel protein digestion and peptide extraction [57-59]. Briefly, proteins were introduced into 12% SDS-PAGE gels using a constant voltage (125 V) to approximately 1 cm into the resolving gel. The proteins were stained by CBB, destained in 30% methanol, 10% glacial acetic acid and 60% distilled H₂O, followed

by excision of stained gel bands. The gel bands were completely destained, treated with 10 mM DTT at 56°C for 45 min, 55 mM IAA at room temperature for 30 min in the dark and digested at 37°C for 16 h with 2 µg porcine trypsin (sequencing grade, Promega). The resultant tryptic peptides were extracted in 40% ACN/0.1% TFA into fresh Protein LoBind® microtubes, dried by vacuum centrifugation, and resuspended in 50 µL 0.1% TFA. The peptides were desalted using oligo™ R3 reversed-phase resin (Cat. # 1133903, Thermo Fisher) and subsequently dried as before.

Glycopeptide enrichment and deglycosylation

Hydrophilic interaction liquid chromatography (HILIC) was employed to enrich for glycopeptides [60-65]. Briefly, the desalted tryptic peptides were passed three times through a p200 pipette tip packed onto a C8 disk (Empore) with PolyHYDROXYETHYL ATM HILIC SPE resin (PolyLC Inc). The microcolumns were washed using 80% (v/v) ACN/1% (v/v) TFA, followed by the elution of the bound glycopeptides in 50 µL 0.1% (v/v) TFA, 50 µL 25 mM NH₄HCO₃ and finally in 50 µL 50% (v/v) ACN. The three fractions were combined, dried by vacuum centrifugation, resuspended in 20 µL 50 mM NH₄HCO₃ and PNGase F (250 U) added. Deglycosylation was performed for 16 h at 37°C prior to desalting on two C18 disks (3M Empore C18 disks, product number 2215) packed in a p200 tip [66].

nLC-MS/MS analysis

Formerly N-glycopeptides were separated by nano-LC-MS/MS coupled to an ESI-LTQ-Orbitrap Velos mass spectrometer (Thermo Fisher Scientific). Mass spectrometric analysis was performed with the following settings: The injection volume was 10 µL, and the total analysis time for each sample was 105 min. The 20 most intense precursors selected from the Orbitrap MS1 full scan (resolving power 120,000 full width at half-maximum (FWHM) @ m/z 200) were isolated and fragmented by collision-induced dissociation (CID) and detected in the dual-pressure linear ion trap with 35 as normalized collision energy (NCE). The dynamic exclusion duration was set to 15 s.

Identification and quantification of proteins and peptides in the formerly N-glycosylated peptide (FNGPs) fraction were performed using MaxQuant v.1.5.3.8 [67] using the label-free quantification (LFQ) algorithm. The MS/MS spectra were searched against the *Macaca mulatta* proteome database in UniProtKB (downloaded, April 2020 with 44,402 entries). Trypsin was set as the proteolytic enzyme with 2 missed cleavages, and the minimum peptide length was set at 7 amino acid residues. The precursor ion and fragment ion mass tolerance were set at 20 ppm and 0.5 Da, respectively while the protein FDR was set at 0.01. Carbamidomethylation of cysteines was set as a fixed modification, while deamidation (N) was set as a variable modification and used for protein

quantification. Localization probability was calculated using the algorithm embedded in MaxQuant [68].

Quantitative mass spectrometry analysis of the total proteome

Proteome-level changes in LLC-MK2 cells treated with or without DTT were evaluated. The total proteins were extracted as previously described, followed by in-gel denaturation and alkylation steps using 10 mM DTT and 55 mM IAA, respectively. The proteins were in gel-digested overnight with sequencing grade trypsin (Promega). The resulting tryptic peptides were desalted using in-house C18 disks, and the peptides submitted to nLC-MS/MS analysis as previously described. Carbamidomethylation of cysteines was set as a fixed modification, while oxidation of methionine and protein N-terminal acetylation were set as the variable modifications, and used for protein quantification. nLC-MS/MS analysis was performed as previously described.

Bioinformatics and data analysis

Statistical analyses were performed in Perseus v.1.6.10.43 [69] and *GraphPad Prism v.8.0* (*GraphPad Software, La Jolla California USA, www.graphpad.com*). Contaminants and proteins identified in the reverse database were excluded before statistical analyses. For both total proteome and deglycoproteome analyses, a cut-off was included to filter for proteins and FNGPs, respectively, which were identified and quantified in duplicates in at least one condition. Regulated proteins between LLC-MK2 cells treated with and without DTT were analyzed by student t-test using Benjamini-Hochberg based FDR (FDR < 0.05) to test the effect of DTT on protein expression. Analysis of FNGPs was performed to evaluate the effect of DTT on 1) N-glycosylation and 2) structural conformational changes, respectively, as described below. Deamidated peptides with the glycosylation motif ((Asn[Xxx]Ser/Thr/Cys) where Xxx ≠ Pro)[70] and localization probabilities ≥ 0.9 (90%) were considered for subsequent analyses. t-test analysis of LFQ values of the protein groups from control and DTT treated cells in the absence of PNGase F treatment was performed using Benjamini-Hochberg based FDR (FDR < 0.05). The total ion intensities of peptides from control cells (treated with or without PNGase F) and DTT treated cells (with or without PNGase F) were evaluated using ANOVA with Tukey's test to correct for multiple comparisons in *GraphPad Prism v.8.0*. Statistical analysis of FNGPs was performed by ANOVA using Benjamini-Hochberg FDR (FDR < 0.05).

Gene ontology analysis

Gene ontology (GO) analysis of the glycoproteins identified and quantified by mass spectrometry analysis was determined using the g:profiler tool [71]. Enriched cellular components, molecular

functions and biological processes were determined by applying a q-value cut-off of 0.05, corrected by Benjamini-Hochberg FDR.

Glycoprotein structure analysis

Selected glycoproteins with pronounced structural changes upon DTT treatment were analyzed to determine their 3D structures and localization of the N-glycosylation sites based on available crystal structures of homologous proteins deposited in the Protein Data Bank (PDB) (<https://www.rcsb.org/>). For this analysis, the Protein Homology/analogY Recognition Engine v.2.0 (Phyre²) [72] was used to model and predict the 3D structure of selected glycoproteins identified by mass spectrometry displaying pronounced structural changes upon DTT treatment. The glycoprotein 3D structures were analyzed for their localization of the mapped N-glycosites and solvent accessible surface area (SASA) of the modified N-glycosylation sites using Pymol v.2.3.4, with dot density set to 4, and a solvent radius of 1.4 Å. To enable 3D structure alignment and comparison with homologous glycoprotein structures already deposited in the PDB, PyMod plugin of PyMOL [73-75] was used for alignment to determine if the positions undergoing pronounced structural modulations upon DTT treatment were conserved between *M. mulutta* and *H. sapiens* and/or *M. musculus* orthologous glycoproteins. For this, previously modelled LLC-MK2 glycoproteins from our study were loaded on PyMod, and a PDB database search using remote BLAST performed using a E-value threshold of 10 was used to retrieve similar 3D-structures in PDB. High scoring PDB structures were downloaded and the 3D structures aligned. The N-glycosites undergoing pronounced conformational changes based on our analysis were mapped on the homologous glycoproteins and illustrated.

In addition, identified N-glycosites undergoing pronounced and subtle structural changes upon DTT treatment were analyzed using Prediction of Order and Disorder by evaluation of NMR data (ODiNPred) [76], a sequence-based protein order/disorder predictor. The evolution option was activated for this analysis.

Results and Discussion

Proof-of-concept and optimization of the limited deglycosylation assay (LDA)

In this study, we introduce the limited deglycosylation assay (LDA) to study conformational changes of glycoproteins on a systems-wide scale using a quantitative label-free analysis of deglycosylated peptides. The experimental workflow for LDA is summarized in **Figure 1**. In step 1, standard

glycoproteins, cells or tissues are subjected to biotic or abiotic stress influencing glycoprotein folding and structure. Proteins are extracted under native conditions using a mild detergent to improve membrane solubilization while avoiding disruption of protein conformation (step 2), before treatment of the still natively folded proteins with PNGase F (step 3). The glycoprotein folding states dictate the susceptibility to PNGase F-mediated removal of their surface exposed N-glycans. Subsequently, proteins are loaded into a SDS-PAGE to remove the detergent and subsequently in-gel digested using trypsin (step 4). Glycosylated peptides are enriched using HILIC (step 5) and subjected to deglycosylation using PNGase F (step 6). Resulting FNGPs are analysed using label-free quantitative mass spectrometry (MS)-based proteomics. The differential ion intensity of FNGPs allow a site-specific mapping of PNGase F accessibility which can be correlated with conformational changes in the 3D structures upon stress.

Initially, the LDA method was optimized using RNase B and Fetuin to establish the optimal detergent concentration, PNGase F-to-glycoprotein ratio, and the reaction time and temperature (**Figure 2**). Bovine RNase B carries high mannose-type glycans at a single N-glycosite, while fetuin has 3 N- [61, 77-79] and 6 O-glycosites [80, 81]. The two glycoproteins were partially deglycosylated under native conditions, while their denatured forms, included as controls, were completely deglycosylated, as illustrated by their differential migration on SDS-PAGE. In our analysis, PNGase F hydrolysed N-glycans more efficiently in 0.4 % NP-40 compared to 1% NP-40 buffer for both RNase B and fetuin, as shown by the intensity of bands corresponding to the completely deglycosylated controls (**Figure 2A**). The impact of the deglycosylation temperature was evaluated at 4°C, 24°C and 37°C in native buffer with 0.4% NP-40, which showed a higher deglycosylation efficiency at higher temperatures (**Figure 2B**). At 37°C, the band corresponding to the fully deglycosylated band was observed, highlighting temperature-dependent changes in glycoprotein structural conformation and subsequent differential sensitivity of PNGase F at 37°C compared to 4°C and 24°C. Higher degree of deglycosylation of RNase B and fetuin at PNGase F:glycoprotein substrate ratio of 1000 U/2.5 µg (corresponding to 400 U/µg) at 4°C for 16 h in 0.4% NP-40 buffer was demonstrated, while decreasing rates of deglycosylation were demonstrated up to 15 U/µg (**Supplementary Figure 2C, 2D**). Incubation time of 16 h in native conditions demonstrated higher deglycosylation levels for both glycoproteins (**Figure 2E, F**). Taken together, native PNGase F-mediated deglycosylation was shown to be most efficient in 0.4 % NP-40 buffer for 16 h incubation at 4°C preserved the native structural forms of the glycoproteins. Thus, these conditions were used for the subsequent LDA experiments with cellular proteins following DTT-based induction of ER stress. PNGase F has low

activity on native glycoproteins [82], while misfolded and/or unfolded/denatured glycoproteins are susceptible to cleavage of N-glycans by endogenous endoglycosidases [83, 84]. Endo H and PNGase A have even less activity on native glycoproteins [83]. Methods to probe the structural conformational states of standard glycoproteins using PNGase F coupled to SDS-PAGE, lectin blotting, mass spectrometry, NMR procedures have previously been demonstrated [82-84]. However, LDA is the first method to study the systems-wide glycoprotein structural modulation in a biological system.

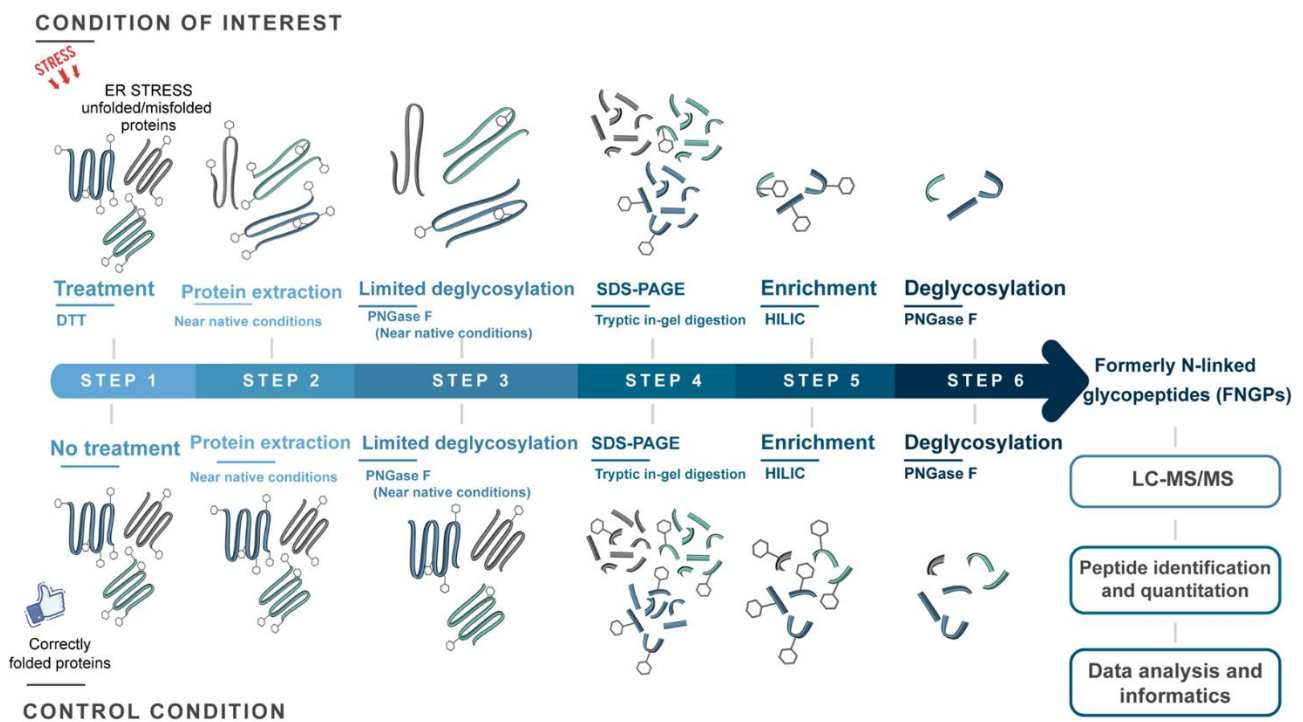


Figure 1. Experimental workflow of the limited deglycosylation assay (LDA). LDA probes glycoprotein conformational changes on a global scale following perturbation-induced structural modulation. Cells treated or not with ER stress inducers are lysed in native conditions and subjected to PNGase F treatment (4°C for 16 h). The resulting proteins are analyzed by GeLC approach using SDS-PAGE and in-gel digestion of the entire lane. The tryptic glycopeptides are enriched using HILIC SPE and subsequently deglycosylated by overnight PNGase F treatment. Quantitative nLC-MS/MS analysis was performed to identify differentially formerly N-glycopeptides (FNGPs).

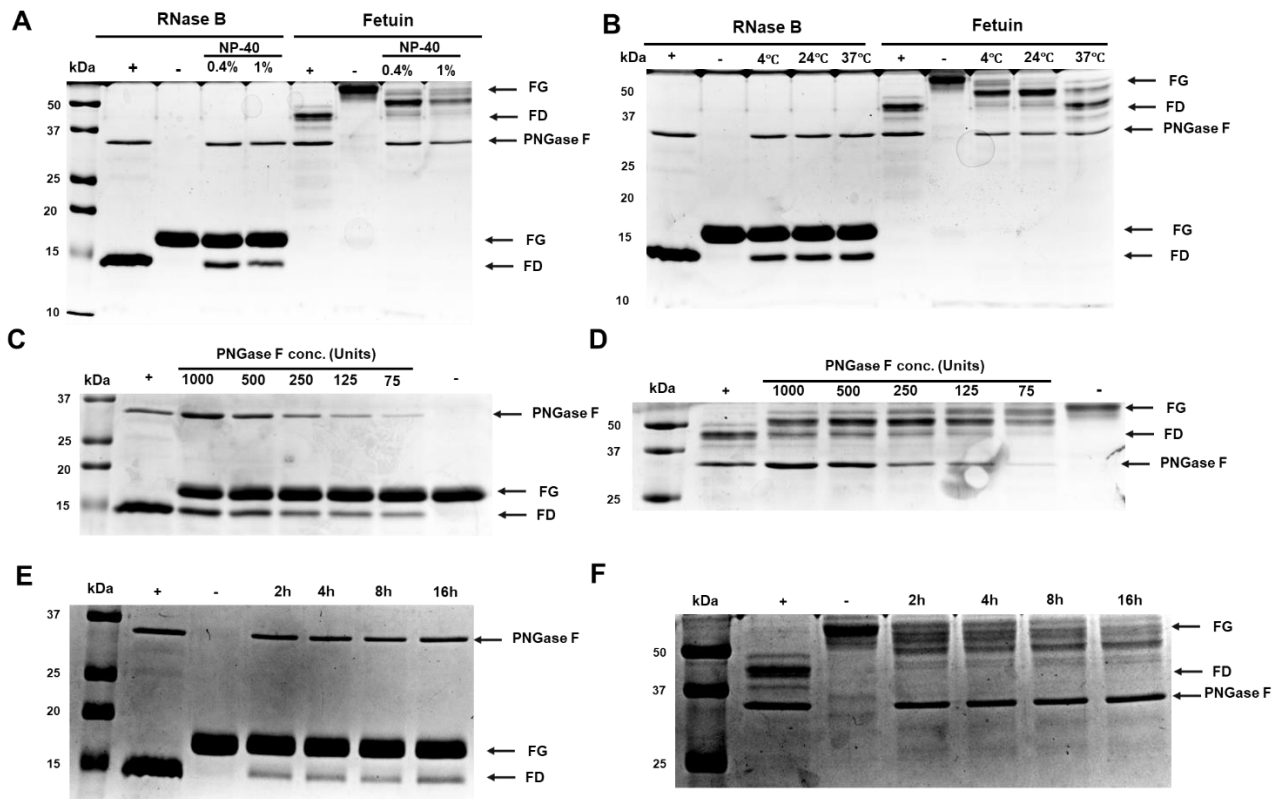
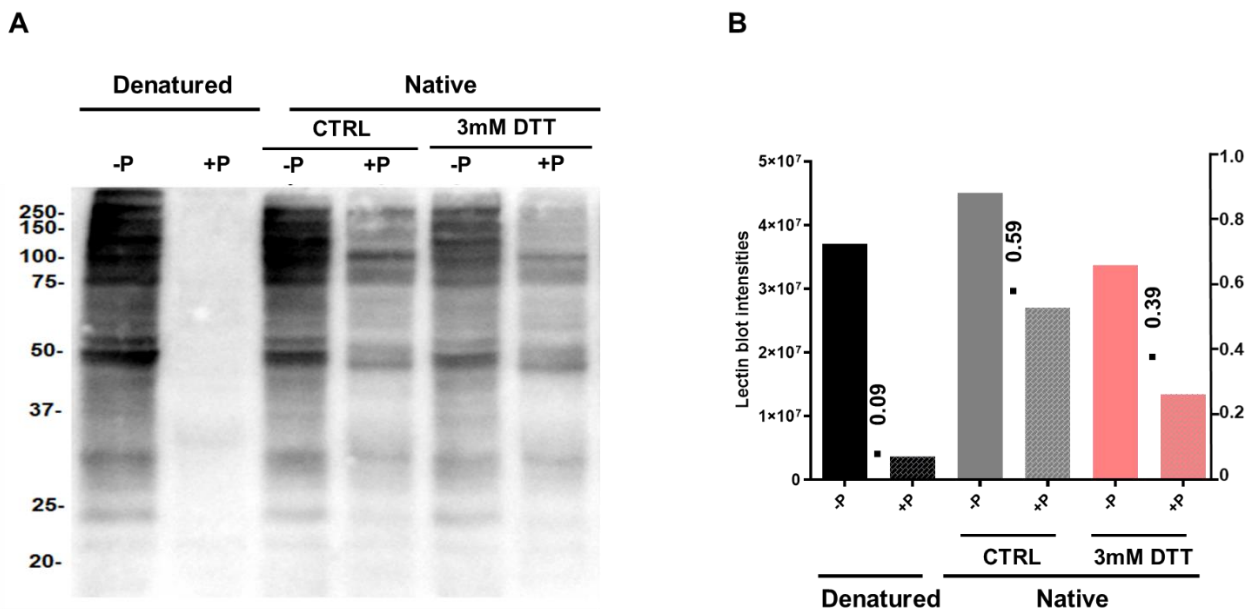


Figure 2. Optimization of the LDA method using RNase B and fetuin. Experimental conditions for optimal limited deglycosylation of bovine RNase B and fetuin by PNGase F were determined. The optimal **A)** NP-40 concentration, **B)** temperature, **C)** and **D)** PNGase F:glycoprotein ratio for RNase B and fetuin, respectively, and, **E)** and **F)** the optimal time for the deglycosylation of RNase B and fetuin, respectively, were assayed by differential migration of the glycoproteins in 15% SDS-PAGE. Fully deglycosylated (+) and fully glycosylated (-) proteins were included for each optimization reaction, denoted by FD and FG, respectively.

Proteins extracted from LLC-MK2 cells treated with or without 3 mM DTT were subjected to limited deglycosylation under native conditions and analyzed using lectin blotting to confirm the optimal conditions previously determined on standard glycoproteins (**Supplementary Figure 1A**). Cells were lysed in native lysis buffer with 0.4% NP-40, and treated overnight with 1000 U PNGase F at 4°C or left untreated. A denatured protein control was included, where equal amount of proteins was initially boiled in glycoprotein denaturing buffer for 10 min at 95°C, cooled on ice prior to overnight incubation with PNGase F as described above. Notably, the complete deglycosylation (0.09) of the denatured fraction was achieved analysing the ratio between lectin blot intensities in the +PNGase F divided by -PNGase F conditions (+PNGase F/-PNGase F = 0.09) (**Supplementary Figure 1B**). Under native conditions, however, limited deglycosylation was observed, where more deglycosylation of proteins from cells treated with 3 mM DTT (+PNGase F/-PNGase F = 0.39) compared to proteins from cells incubated in the absence of DTT (+PNGase F/-PNGase F = 0.59) (**Supplementary Figure 1B**).

These data support that PNGase F is able to differentially remove N-glycans from proteins with different structural conformations depending on its accessibility to the N-glycans on the natively folded glycoproteins. As shown for the denaturing conditions where the proteins are completely un- or misfolded (**Supplementary Figure 1**), all the N-glycans are hydrolyzed by PNGase F, illustrating their improved susceptibility to hydrolysis by the enzyme compared to the partial hydrolysis under native conformations. LDA on DTT treated cells revealed a higher deglycosylation rate compared to control cells grown in the absence of DTT, revealing conformation-based differential accessibility of PNGase F to N-glycans (**Supplementary Figure 1B**). Incubation of cells in DTT treated culture medium prevents the formation of disulphide bond in nascent polypeptide chains in the ER as well as reduction of disulphide bonds in already folded proteins, which results in conformational modulation and subsequent accumulation of misfolded/unfolded proteins in the ER [85, 86].

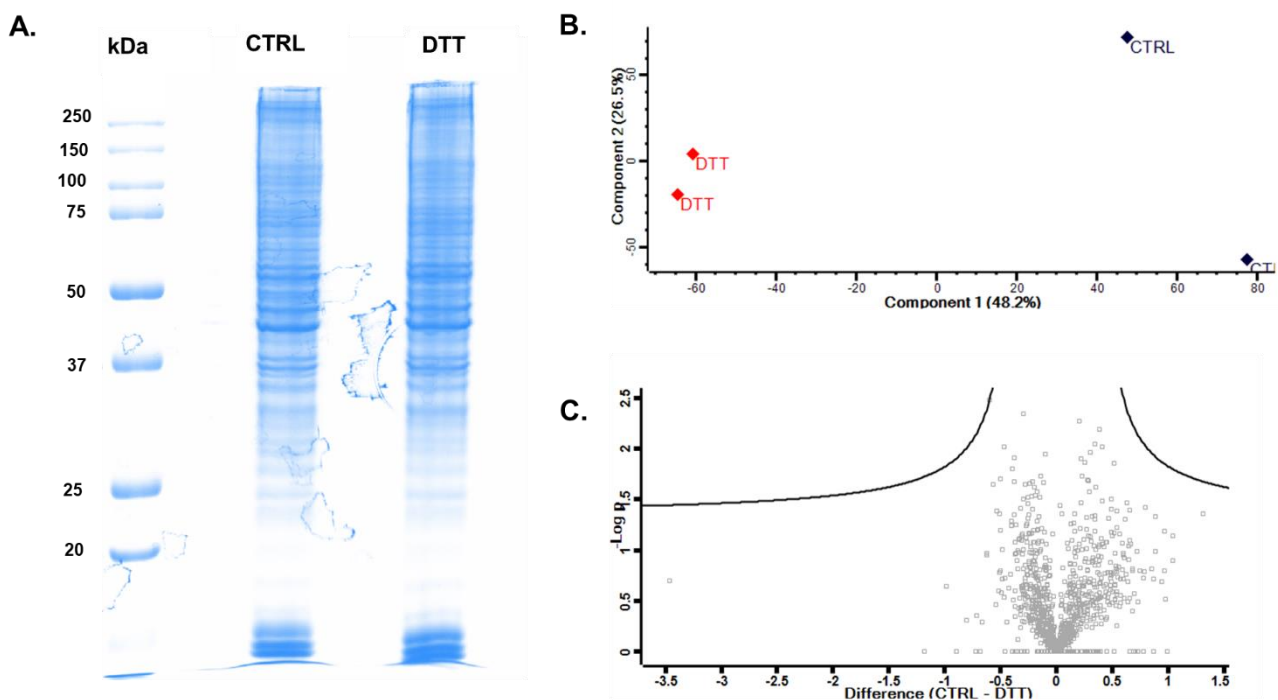


Supplementary Figure 1. Optimized LDA applied to LLC-MK2 cellular glycoproteins. A) Lectin blotting using Con A on LLC-MK2 cells following LDA using optimized conditions previously determined on standard glycoproteins. **B)** Intensity and ratios between +PNGase F/-PNGase F calculated using ImageLab are shown. -P = negative control not treated with PNGase F; +P = PNGase F treated.

Proteome-wide analysis of LLC-MK2 cells treated with or without 3 mM DTT was performed prior to LDA to evaluate the proteome-wide modulation upon incubation with DTT. Protein band patterns were initially visualized by SDS-PAGE (**Supplementary Figure 2A**), followed by an in-depth bottom-up label-free mass spectrometry-based proteomics approach. No notable differences between control and DTT treated cells were detected through SDS-PAGE, as indicated by the protein bands intensity ratios between DTT and CTRL determined to be 1.06. Quantitative proteome data showed

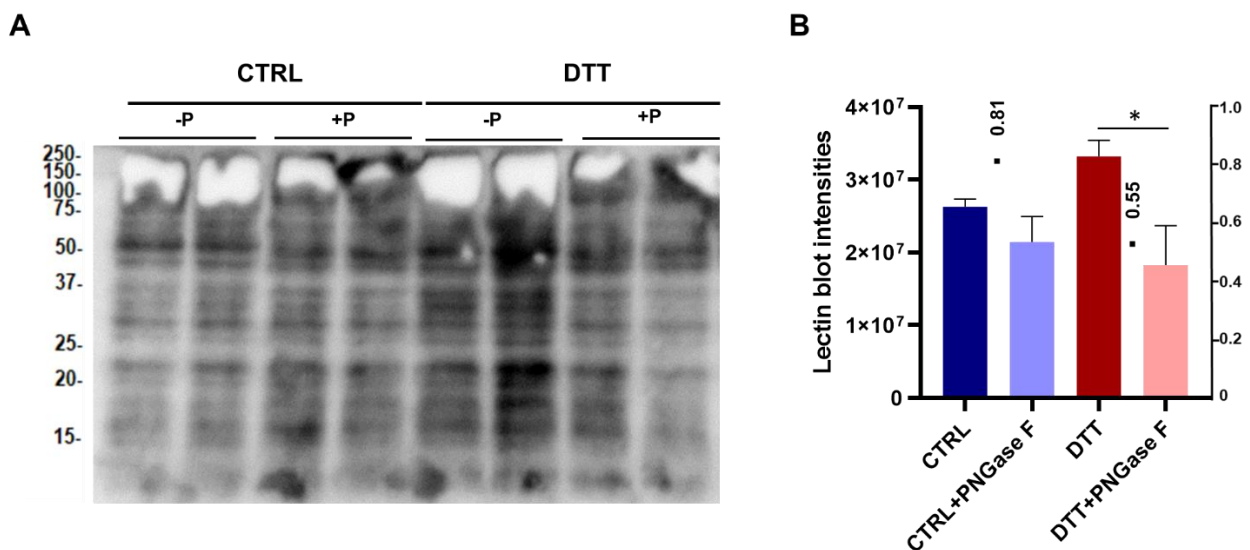
a separation of the control and DTT treated fractions by principal component analysis (PCA) (**Supplementary Figure 2B**). However, paired t-test analysis using Benjamini-Hochberg based FDR at an FDR < 0.05 showed that upon 3 mM DTT treatment for 30 min, the regulation in protein abundance at the protein level was not statistically significant (**Supplementary Figure 2C**). Longer incubation periods of 18 h with 2 mM DTT have been shown to cause clear UPR activation in HeLa cells along with thapsigargin and tunicamycin [22]. In this study, LLC-MK2 cells were incubated for 30 min in the presence of 3 mM DTT, a short time to observe complete UPR activation, but sufficient time to analyze glycoprotein conformation changes prior to protein level abundance changes due to DTT induced ER stress.

A total of 1540 proteins were identified and quantified in this study using label-free mass spectrometry (**Supplementary Table 1**). Of these, 1053 proteins were identified in 2 replicates in at least one condition, and they were used for subsequent analyses. A total of 9 and 8 proteins were identified exclusively in control and DTT fractions, respectively (**Supplementary Table 1**). Analysis of the total quantitative proteome data revealed subtle increase in ER stress associated proteins such as the chaperones HSP90 β 1 protein (I0FUR4/endoplamin) and 78 kDa glucose-regulated protein (F7C3R1/HSPA5), calnexin (H9ES13/CANX), calreticulin (G7NLC3/CALR), and protein disulphide isomerases (A0A5F8ABB1/PDIA4; F6W4X0/PDIA3; and A0A1D5RKD4/P4HB) (**Supplementary Table 1**), suggesting the early onset of UPR.



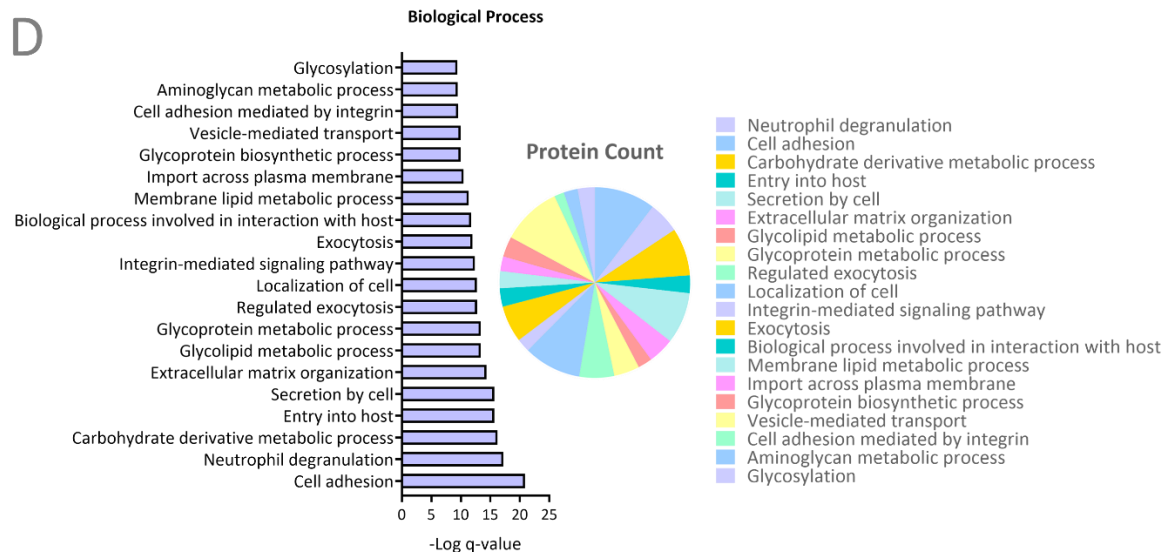
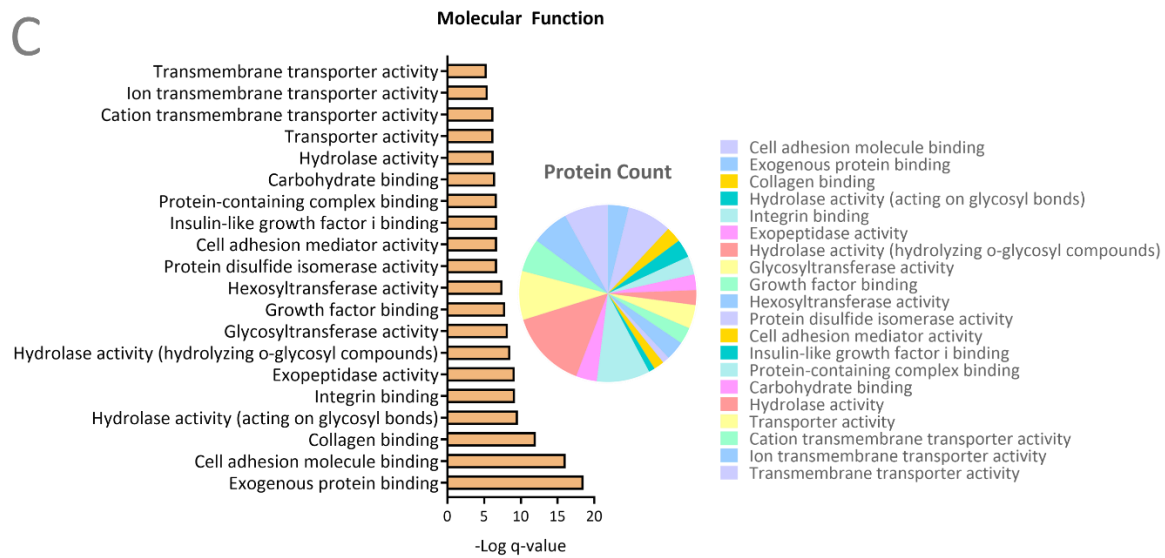
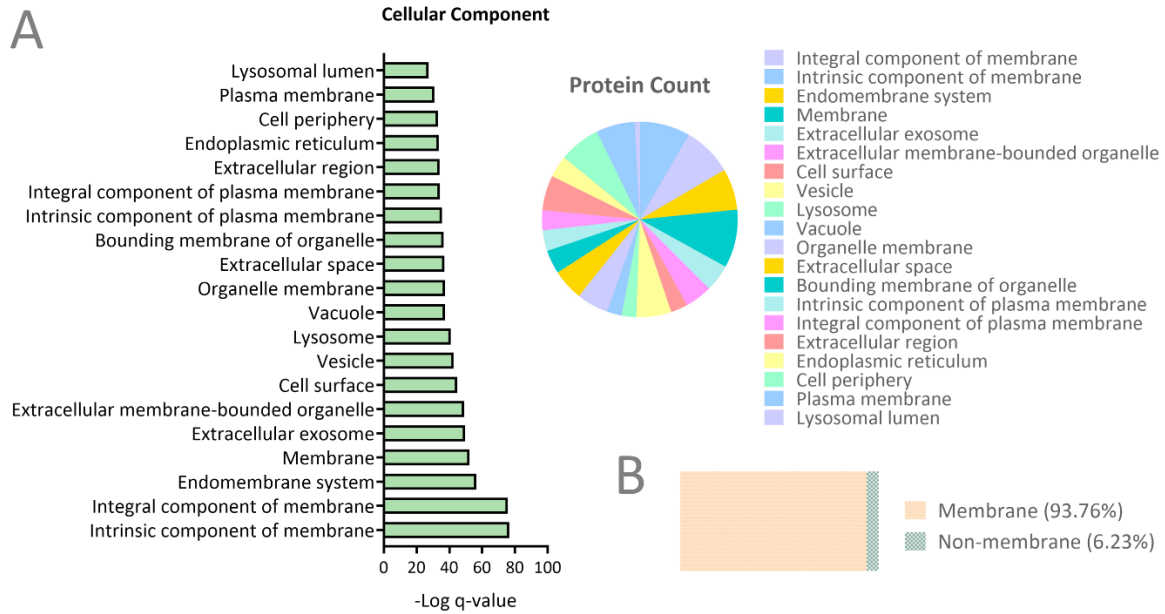
Supplementary Figure 2. Protein expression analysis of LLC-MK2 cells incubated in medium supplemented with or without 3 mM DTT for 30 min. (A) The protein expression profiles were visualized by SDS-PAGE, and the band intensities analyzed using ImageLab. (B) The effect of DTT on the proteome was visualized using PCA, and the levels of regulation were visualized using (C) Volcano plot calculated at a FDR < 0.05.

LDA was performed using the optimized conditions and the deglycosylation rate was monitored using Con A blotting before trypsin digestion, glycopeptide enrichment and mass spectrometry analysis. Lectin blot intensities corroborated that more N-glycans were cleaved off in the DTT treated fractions compared to control fractions (**Supplementary Figure 3A**). A statistically significant reduction in levels of glycosylation between DTT+PNGase F and DTT only fractions (ratio = 0.55) was seen using one-way ANOVA using Tukeys correction of multiple comparisons ($p < 0.05$) (**Supplementary Figure 3B**). In addition, higher levels of glycosylation were observed in DTT treated fractions compared to control fractions in the absence of PNGase F treatment (**Supplementary Figure 3B**). It has been shown that the prevention of disulphide bond formation and subsequent protein misfolding by treatment of cells with mild levels of DTT leads to complete glycosylation of N-sequons that, in native conformations, are missed or undergo variable levels of glycosylation [87, 88].



Supplementary Figure 3. Lectin blot analysis of LLC-MK2 cells incubated with or without 3 mM DTT for 30 min followed by deglycosylation under native conditions. A) ConA lectin blot following LDA. B) The combined intensities for each condition are summarized, and the ratios between +PNGase F/-PNGase F for both control and DTT treated fractions. Results are displayed as mean \pm SEM; * $p < 0.05$.

Analysis of the FNGPs by MS/MS following limited deglycosylation and HILIC enrichment identified a total of 1403 deamidated peptides, of which 1240 were identified in 2 replicates in at least one condition, representing 571 unique source glycoproteins (**Supplementary Table 2; Figure 2A**). To ensure high confidence of the deamidated peptides identified and quantified in this study, peptides with a localization probability ≥ 0.9 (90 %) were selected, resulting in 1145 deglycopeptides. Of these, 1001 peptides, corresponding to 433 unique source glycoproteins, had one or more N-glycosylation sequence motifs, and were considered for further analysis. Gene ontology analysis of the identified glycoproteins is illustrated in **Supplementary Figure 4**. Plasma membrane, organelle membrane and extracellular membrane-bounded organelles terms were enriched as cellular components (**Supplementary Fig. 4A**). Indeed, 93.7% of the total identified glycoproteins were associated to membrane showing the robustness of the N-linked glycopeptide enrichment and site analysis (**Supplementary Figure 4B**). Binding of cell adhesion molecules, transmembrane transporters and glycosyltransferases activity were enriched as molecular functions (**Supplementary Figure 4C**). The identified glycoproteins were associated to cell adhesion, integrin-mediated cell signaling, glycosylation, secretion and vesicle-mediated transport as biological processes (**Supplementary Figure 4D**).



Supplementary Figure 4. Gene ontology analysis of LLC-MK2 deglycoproteome following LDA. Top 20 Gene Ontology terms of **A**) Cellular components, **B**) percentage of proteins that localize to the membrane. **C**) Molecular functions and **D**) Biological processes are illustrated. Only terms with q-value <0.05 were considered (Benjamini-Hochberg).

Statistical analysis by ANOVA at an FDR < 0.05 showed that 123 peptides were statistically regulated between the four fractions (**Figure 2A**), representing 101 unique source glycoproteins. The enrichment step using HILIC in our current study is paramount to increase the coverage of glycopeptides, as revealed by Venn diagram analysis showing that only 45 glycoproteins were identified and quantified in common between the proteome and deglycoproteome datasets (**Figure 2B**). It should also be noted that measuring protein levels between the two conditions is also important to exclude any protein expression-dependent changes that can be confounding factors for determining differential abundance of FNGPs. In this study, a short stimulation with DTT did not influence the protein levels (**Supplementary Figure 2C**).

Deglycoproteome analysis by PCA and Euclidean distance showed that presence and absence of PNGase F treatment during the limited deglycosylation step 3 (**Figure 1**) accounted for the majority of the variation in the dataset (**Figure 2C and D**). A clear separation into 2 major clusters was observed based on the first principal component, which accounted for 41.5% of the total variation in the dataset (**Figure 2C**). Separation between DTT treated and control fractions in the absence of PNGase F treatment was observed based on the second principal component, representing 13.6% of total variation in the dataset, which highlighted the potential effect of DTT on LLC-MK2 cellular glycosylation (**Figure 2C**). Statistical analysis using ANOVA at an FDR < 0.05 identified 123 regulated FNGPs. The regulated FNGPs formed three major sub-clusters which included: i) glycopeptides whose abundances increased following DTT incubation compared to the control fractions in the absence of PNGase F treatment, and subsequently were completely deglycosylated in DTT fractions compared to their PNGase F untreated counterparts; ii) glycopeptides whose structural conformations in native and mis/unfolded conformations had equal accessibility by PNGase F, and iii) glycopeptides whose PNGase F accessibility was more following DTT incubation compared to the -DTT control condition. The structure of selected proteins in subset iii) was further analyzed and discussed.

Total ion intensities of FNGPs with the N-glycosylation sequence motif revealed statistically significant differential deglycosylation by PNGase F, showing that more N-glycans were accessible to PNGase F cleavage in the DTT treated fractions compared to native conditions (**Figure 2E**).

A total of 43 deamidated peptides were identified and quantified in subset iii), which included FNGPs missing exclusively in the DTT treated fractions incubated overnight in the presence of PNGase F. The corresponding glycoproteins were considered to have undergone pronounced structural modulation upon DTT treatment, exposing the N-glycans to PNGase F, leading to their complete cleavage (**Supplementary Table 2**). Three glycoproteins undergoing pronounced and/or subtle conformational modulation upon DTT treatment, with 1, 2 and 4 mapped N-glycosites were chosen for further discussion following 3D structural analysis.

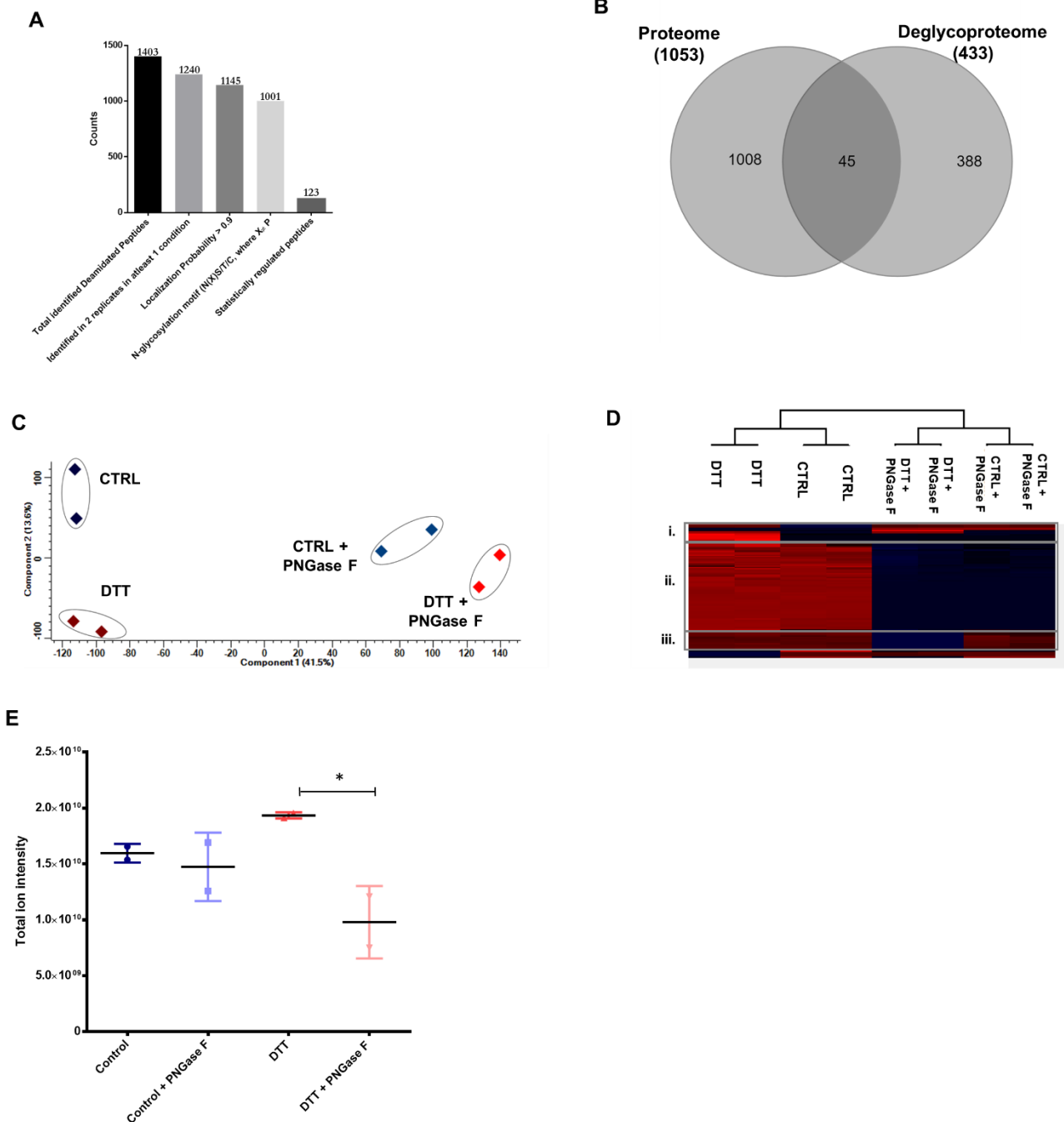


Figure 3. Limited deglycosylation of LLC-MK2 cells treated with or without 3 mM DTT. The mass spectrometry analysis showing **A)** number of identified and quantified deaminated peptides, **B)** Venn diagram showing shared proteins from proteome and deglycoproteome analysis, **C)** the PCA **D)** heatmap of deaminated peptides, **E)** total ion intensities of the deaminated deglycopeptides from control, control + PNGase F, DTT, and DTT + PNGase F fractions.

One N-glycosite (N37) of Ephrin-A5 (F7GZC7/EFNA5), a membrane GPI-anchored protein, was identified with profound structural modulation. Analysis of the global breast cancer xenograft tissue proteome identified the human homolog to this protein to be glycosylated [89], with Hex₅₋₆ HexNAc₂

glycans at N37 reported in Glyconect. The human homolog ephrin-A5 ectodomain crystal structure has been determined by X-ray crystallography [90]. Using Phyre2, the 3D structure of the identified LLC-MK2 cellular ephrin-A5 was modelled, revealing that N37 is located on a loop, 3 amino acid residues away from a beta sheet (**Figure 4A**). Under native conditions, this site is completely shielded from PNGase F accessibility, as the ratio between CTRL + PNGase F and CTRL is 1.05. The SASA of ephrin-A5 N37 was determined to be 34.3 Å² using PyMol, indicating that this site has a low solvent-exposed surface area (**Figure 4B**). Structure alignments of human ephrin-A5 (4LOP chain B) [91] and *Mus musculus* ephrinA5 (1SW chain A) [92] showed the conservation of N37 on a loop across species (**Figure 4C**), suggesting possible structural modulation of this protein in different organisms upon ER stress. This glycosite is shared within ephrin-A ligands. It has been shown that ephrin-A1 ligand glycosylation interferes with ephrin-A2 receptor binding, internalization and degradation affecting the downstream signalling pathways. The deglycosylated form of ephrin-A1 does not inhibit cancer cell migration and impairs ligand secretion showing the importance of ephrin-A ligands glycosylation [93]. Analysis of ordered and disordered localization of the N-glycosite using ODiNPred showed that at position N37, the probability of the N-glycosite to be located in a disordered region was 0.0342 (**Figure 4D**). All N37 neighbouring amino acid residues had low probabilities of disorder, in agreement with findings on N-glycosites of human proteins [94].

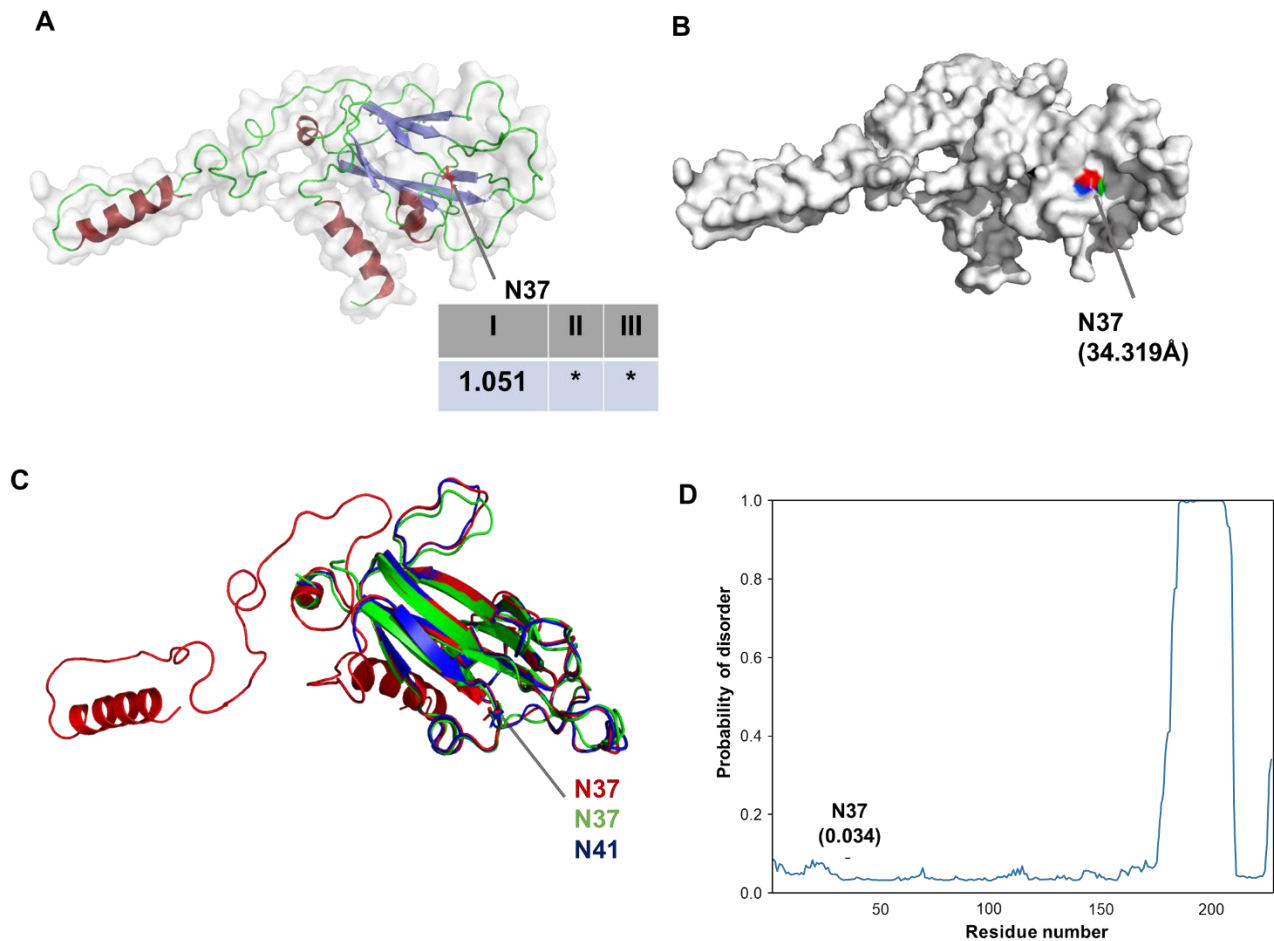


Figure 4. 3D structure of ephrin-A5 (F7GZC7) showing the mapped N-glycosylation site. **A)** Cartoon representation of the modelled 3D structure of ephrin-A5 illustrating the ratios of the MS/MS intensities of FNGPs between PNGase F-treated and -untreated cell extracts. α -helix are in firebrick red, β -sheet in slate blue, and loops in green. **I** represents the intensity ratios of CTRL+PNGase F/CTRL, **II** represents the ratio between DTT+PNGase F/DTT, and **III** represents the ratio of **II**/I, * represents ratios with DTT+PNGase F equal to 0, indicating deglycopeptides exclusively missing in DTT+PNGase F fractions. **B)** Surface representation showing the SASA of the N-glycosite calculated using PyMol, where the warmer and cooler colors represent N-glycosites with more and less surface exposure, respectively. The SASA of the mapped N-glycosite (in Å) is shown in parenthesis. **C)** 3D structure alignment of modelled *M. mulatta* (red) ephrin-A5 and determined 3D structures of homologs from *H. sapiens* (green) and *M. musculus* (blue) illustrating the conserved localization of the N-glycosite undergoing pronounced structural conformational modulation upon DTT treatment. The N-glycosites are represented using sticks to show the side chains of the glycosylated Asn. **D)** Localization probabilities of ephrin-5A amino acids calculated using ODiNPred, with the mapped N-glycosite (N37) highlighted.

Two of the proteins identified in subset iii) included proteins involved in the protein glycosylation machinery; hexosyltransferase (F6UCJ0/ B3GALNT1) and polypeptide N-acetylgalactosaminyltransferase (A0A1D5QY41/GALNT10). These membrane enzymes reside in the Golgi apparatus. GALNT10 transfers N-acetylgalactosamine from UDP-N-acetyl- α -D-galactosamine to proteins in the Golgi. This enzyme has 6 putative N-glycosites and 13 cysteines. Three N-

glycosylation sites have been mapped on the human homolog of N-acetylgalactosaminyltransferase (Q86SR1/ GALNT10) at positions N124, N146 and N593 [95], glycosites also conserved in *M. mulatta* GALNT10. In addition, the human homolog has 5 disulphide bridges at positions Cys135 ↔ Cys365, Cys356 ↔ Cys432, Cys471 ↔ Cys488, Cys523 ↔ Cys538 and Cys563 ↔ Cys578 [95] sites that are also conserved in *M. mulatta* GALNT10, suggesting close structural similarity between human and *M. mulatta* GALNT10 based on primary amino acid sequence similarities. In our analysis, we identified two FNGPs at positions N128 and N575 (**Figure 5A**), corresponding to N146 and N593, respectively, in the human homolog. At position N128, the N-glycosite was determined to be exclusively missing in the DTT and PNGase F treated fraction, suggesting a pronounced structural modulation at this glycosite compared to the control fractions. Three-dimensional analysis of the modelled protein structure showed that the N-glycosite at this position is located on a loop-alpha helix interface (**Figure 5A**). The N-glycosite at N575 showed that this site underwent subtle structural changes, allowing PNGase F to cleave N-glycans at a higher rate upon DTT treatment compared to control fractions at a ratio of 0.48 (**Figure 5A**). This site is also located on a loop proximal to a beta sheet (**Figure 5A**). The SASA of N128 and N575 were 60.4 Å² and 47.5 Å², respectively (**Figure 5B**), illustrating that under native conditions, N575 is more occluded compared to N128, corroborated by limited deglycosylation assay (i) (**Figure 5A**).

The intensity ratio of the deglycopeptides at N128 and N575 between CTRL + PNGase F and CTRL is 0.82 and 1.14, respectively (**Supplementary Table 2**), corroborating the higher accessibility of N128 by PNGase F cleavage under native conditions compared to N575. N575 becomes more accessible to PNGase F after incubation with DTT, as indicated by a relatively high ratio of 0.55 between the DTT + PNGase F and DTT only fractions. The ratio between DTT and control fractions is 0.49, confirming that under DTT conditions, N-glycans at this site are more prone to undergo PNGase F cleavage compared to non-DTT treated. The 3D structures of the human homolog of GALNT10 (2D7I) and *M. musculus* (1XHB) have been determined [95, 96], and their structures were aligned with the modelled *M. mulatta* GALNT10 glycoprotein, highlighting structural conservation between *M. mulatta* and *H. sapiens* N128 (**Figure 5C**). In *M. musculus*, this N-glycosite is not conserved. Our calculations showed low probabilities (0.077 and 0.079, respectively) for N128 and N575 to localize in disordered regions of the proteins (**Figure 5D**). In addition, the neighbouring amino acids for the two N-glycosites also have low probabilities predictions to be located within disordered regions.

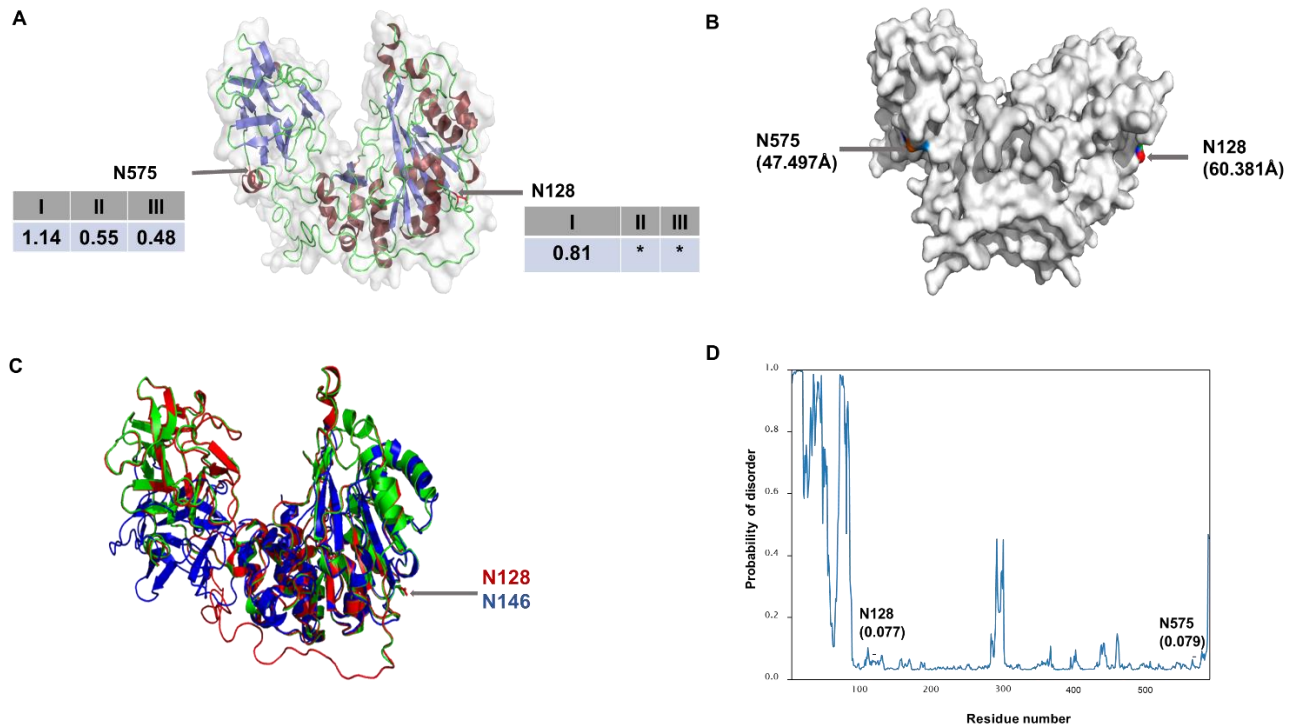


Figure 5. 3D structure of GALNT10 showing N128 and N575 N-glycosylation sites undergoing structural modulation upon DTT treatment. **A)** Cartoon representation of modelled 3D structure of *M. mulatta* N-acetylgalactosaminyltransferase illustrating the ratios of the intensities of FNGPs upon LDA; α -helix are in firebrick red, β -sheet in slate blue, and loops in green. **I** represents the intensity ratios of CTRL+PNGase F/CTRL, **II** represents the ratio between DTT+PNGase F/DTT, **III** represents the ratio of **II/I**, * represents ratios whose DTT+PNGase F/DTT is 0, to denote FNGPs exclusively missing in DTT+PNGase F fractions. **B)** Surface representation showing the SASA of the N-glycosites calculated using PyMol. **C)** 3D structure alignment of modelled *M. mulatta* (red) GALNT10 and determined 3D structures of homologs from *H. sapiens* (green) and *M. musculus* (blue) illustrating the conserved localization of the N-glycosite undergoing pronounced structural conformational modulation upon DTT treatment. The N-glycosites are represented using sticks to show the side chains of the glycosylated Asn. **D)** Localization probabilities of the GALNT amino acids calculated using ODINPred are shown. The probabilities for the N-glycosites to fall into disordered regions have been indicated.

The glycoprotein A0A5F8ABK3 was identified with 4 FNGPs, of which N278 was exclusively missing in the DTT+PNGase F treated fraction. The human homolog of the poliovirus receptor (PVR, P15151/CD155) is a membrane glycoprotein with three immunoglobulin-like domains in the extracellular N-terminal tail. Human PVR was originally identified as poliovirus receptor, in which the V-type domain is necessary and sufficient for viral binding and internalization [97, 98]. Moreover, PVR binds vitronectin [99]. The human homolog has 10 N-glycosylation sites at N105, N120, N188, N218, N237, N278, N307, N313, N352, and N360. In our analysis, we identified 4 FNGPs at N120, N278, N307, and N313 (**Supplementary Table 2**). 3D analysis of the modelled *M. mulatta* PVR showed that 3 of the 4 N-glycosites were located in loops, with the exception of N120 located on a beta sheet (**Figure**

6A). The N-glycan of the human PVR glycoprotein at N120 was previously reported in human urine samples following MS/MS analysis of intact glycopeptides, and determined to be a complex-type glycan (Hex₅HexNAc₄) [100]. N120 was more occluded compared N-glycosites within loops under both DTT treated and untreated conditions as shown by the ratios between CTRL+PNGase F/CTRL, DTT+PNGase F/DTT. Our analysis showed that N278 had the lowest SASA value among the sites located on loops (48.4 Å²) compared to N307 and N313 sites with 56.4 Å² and 63.9 Å², respectively (**Figure 6B**), but N278 were more susceptible to PNGase F cleavage under native conditions (**Figure 6A**). Upon DTT treatment, the N278 was completely deglycosylated, illustrate that this site became completely accessible to cleavage by PNGase. Alignment of the human PVR showed that this site is located on a beta sheet (**Figure 6C**). The glycosylation of N150 and N120 is not essential for viral replication and tissue tropism [101]. Removal of 4 glycosylation sites in the two outermost V and C-like domains did not influence binding to vitronectin [99]. The probabilities of the N-glycosites to fall within disordered regions were determined to be 0.0359, 0.0384, 0.0343 and 0.0329 for N120, N278, N307 and N313, respectively (**Figure 6D**).

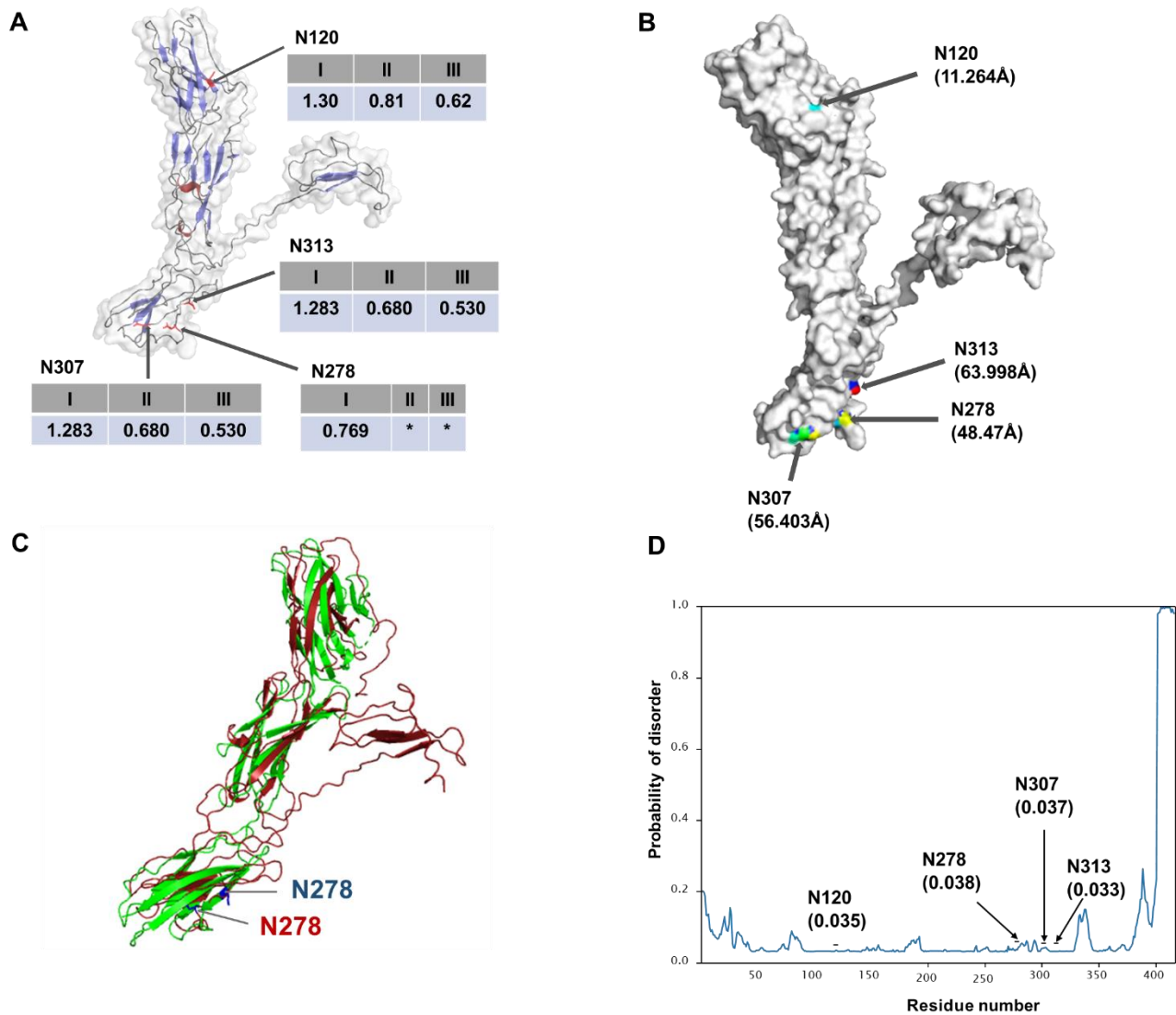
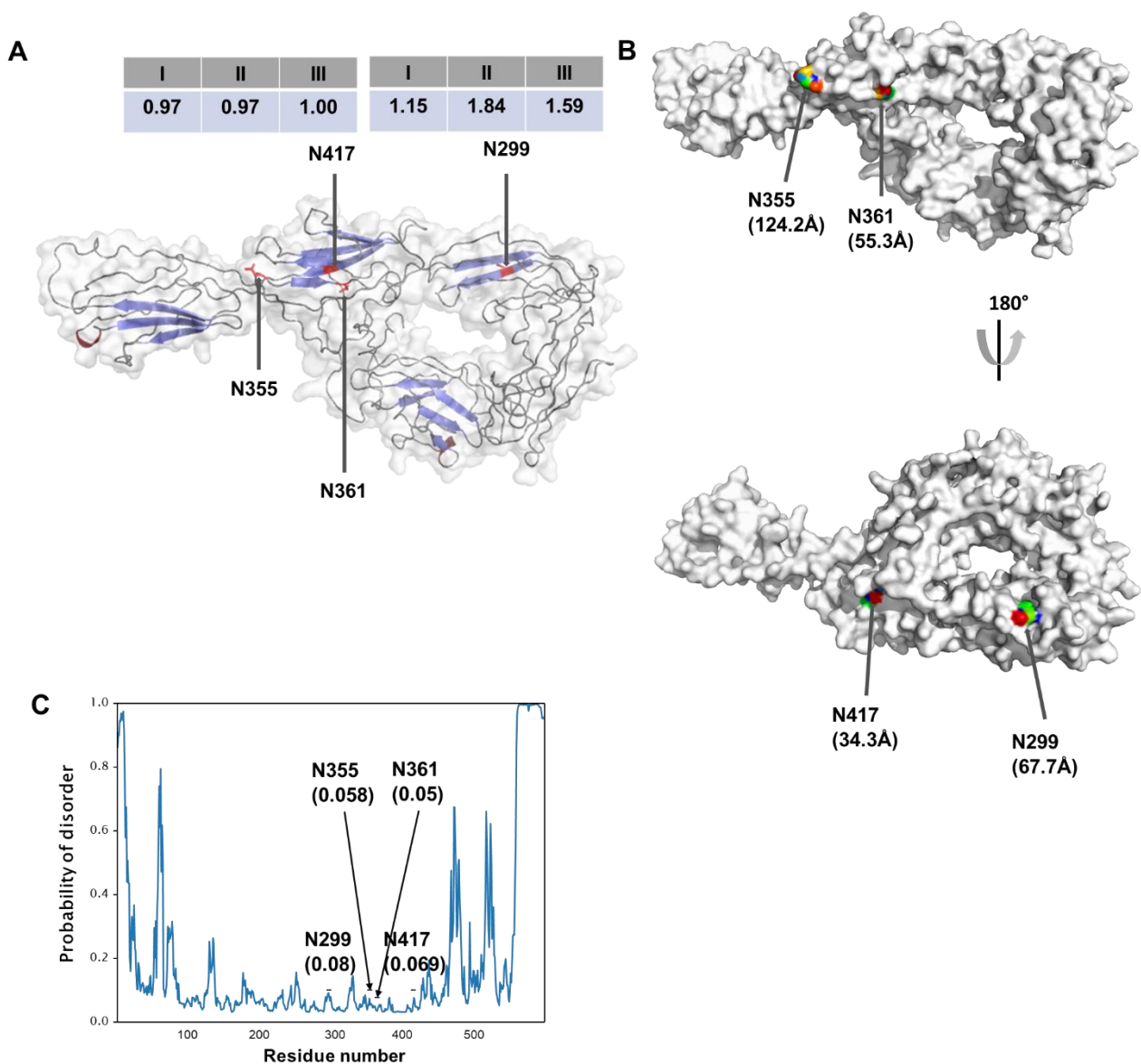


Figure 6. 3D structure of the uncharacterized protein (A0A5F8ABK3/PVR) showing N-glycosites of interest. **A)** Cartoon representation of PVR illustrating the ratios of the intensities of FNGPs. α -helix are in red, β -sheet in blue, and loops in grey. The N-glycosites are in red, and sticks used to represent the side chains. I represents the intensity ratios of CTRL+PNGase F/CTRL, II represents the ratio between DTT+PNGase F/DTT, and III represents the ratio of II/I, * represents ratios whose DTT+PNGase F is 0, representing deglycopeptides exclusively missing in DTT + PNGase F fractions. **B)** Surface representation showing the SASA of the N-glycosite calculated using PyMol, where the warmer and cooler colors represent sites with more and less exposure, respectively. The SASA (in Å) are shown in parenthesis. **C)** Structure alignment of the modelled *M. mulatta* PVR and the determined structure of the human homolog (QARQ) deposited in PDB. N278 is indicated on both proteins. **D)** Localization probabilities of the PVR amino acids calculated using ODINPred. N-glycosites of interest are highlighted.

FNGPs were also identified and quantified in our analysis showing no changes in N-glycosite exposure (and by inference no structural changes in those regions) between CTRL and DTT treated fractions. An example, the basal cell adhesion molecule (A0A1D5Q3L5/BCAM) which was identified with 4 FNGPs; including 2 FNGPs located on beta sheets (N299 and N417), and 2 FNGPs located on loops (N355 and N361) (**Supplementary Figure 4**). The human homolog has been mapped with 5

putative N-glycosites (N321, N377, N383, N419, N439). Human BCAM is known to be involved in adhesion of cells and extracellular matrix components such as laminin, and has been shown to be a receptor of laminin in malignant tumors [102]. The 3D-structure of the modelled *M. mulatta* BCAM revealed that N-glycosites not affected by DTT treatment are located in beta sheets (Supplementary Figure 5A), as indicated by their low SASA (67.8 Å and 34.4 Å, respectively) (Supplementary Figure 5B). N355 and N361 glycosites were affected by DTT treatment (Supplementary Figure 5A). These sites are located in loops as indicated by structural modelling (Supplementary Figure 5A), and analysis of their localization probabilities indicate that they are located in areas with low probabilities of disorder (Supplementary Figure 5C).



Supplementary Figure 5. 3D structure of basal cell adhesion molecule (Lutheran blood group) (A0A1D5Q3L5/BCAM) showing identified N-glycosites. A) Cartoon representation of PVR illustrating the ratios of the intensities of FNGPs. α -helix are in red, β -sheet in blue, and loops in grey. The N-glycosites are

represented in red, and sticks represent the side chains. **I** represents the intensity ratios of CTRL+PNGase F/CTRL, **II** represents the ratio between DTT+PNGase F/DTT, and **III** represents the ratio of **II/I**, * represents ratios whose DTT+PNGase F is 0, representing deglycopeptides exclusively missing in DTT + PNGase F fractions. **B**) Surface representation showing the SASA of the N-glycosite calculated using PyMol, where the warmer and cooler colors represent sites with more and less exposure, respectively. The SASA (in Å²) are shown in parenthesis. **C**) Localization probabilities of the BCAM amino acids calculated using ODINPred. The N-glycosites of interest are highlighted.

The presence of specific PTMs in structural defined motifs has been shown for phosphorylated residues located within unstructured disordered regions both in eukaryotes and prokaryotes [103]. The phosphorylation of these disordered regions allows for enhanced and stabilized protein-ligand binding [104]. Protein O-glycosylation has been mapped also to intrinsically disordered regions protecting specific domains against proteolysis [105, 106]. N-glycosylation has been shown to impact glycoprotein 3D structure and folding through the ER-resident quality control system and thereby is directly involved in dictating the fate of glycoproteins [107, 108]. Thermodynamic and experimental analyses have shown that N-glycosylation often stabilize the glycoprotein structure [109, 110]. N-glycosites are found mainly in ordered regions; amino acid residues surrounding N-sequons are relatively poor in hydrophobic and bulky residues known to promote disordered regions [94]. Proteome-wide structural analysis of N-glycosites from human, mouse, fly, plant and yeast showed that a high proportion of N-glycosites are localized within a turn/loop and are solvent-exposed [111-113]. Approximately 80% of all N-glycosites of the human proteome are located in loop/turn and this pattern is consistent in other eukaryotes [113]. Our data shows that the highly exposed N-glycosites located in turn/loops are more prone to undergo conformational changes upon stress induction compared to sites within alpha helix or beta sheet structures. Moreover, our analysis confirms previous findings showing N-glycosites located in ordered regions. The presence, abundance and occupancy of N-glycosylation enriches the structural features of a glycoprotein that cannot be derived solely based on the primary protein sequence thus conferring a large diversity of functions. The applicability of this method to studying changes in glycoprotein conformation during stress induction can readily be used to study glycoprotein structural changes under other physiological and pathological conditions.

Conclusion

Here we describe a new systems-wide method termed LDA that probes the structural changes of glycoproteins under controlled conditions. LDA explores the differential accessibility of PNGase F to

N-glycosites depending on the folding state of a glycoprotein and uses MS-based proteomics to quantify altered abundance in FNGPs. We developed and validated the method using standard glycoproteins and then showed the potential of the method by evaluating the conformational changes of complex mixtures of glycoproteins in LLC-MK2 cells treated with DTT. Strikingly, the N-glycosites undergoing the highest conformational changes under ER stress induction are located on loops, corroborating findings from previous studies using limited proteolysis. LDA can be applied to study the glyco-proteostasis of different cells or tissues in health and disease, to explore the system-wide modulation of glycoprotein structures in a complex cellular milieu. Since this method is based on MS readout, the detection of most abundant glycoproteins is favoured and an average conformation is detected when multiple glycoprotein folding states co-exist. Moreover, the different accessibility of PNGase F to a specific glycosylation site could be due to differential site occupancy, aggregation, protein-protein interaction, small molecule binding and/or conformational changes. These aspects can be seen as a limitation since it requires specific information on the origin of the changes and a strength since it can be applied to several conditions. The introduction of intact glycopeptide analysis, different enrichment methods, advanced MS-based proteomics tools and combination with other structural approaches will enhance the depth of this method and is thus a logical next step in further method development. We believe that the presented LDA method will serve as a platform to develop other limited enzymatic assays to probe conformational changes based on PTM analysis. Overall, the glycoprotein conformational features deduced from the LDA correlate with those deriving from other biophysical and spectroscopic techniques. LDA will thus contribute to the diverse toolbox required to better study glycoprotein structure and function.

Conflict of interest:

The authors declare no conflict of interest.

Acknowledgments:

Prof. Maria Julia Manso Alves, Prof. Walter Colli and Célia Aparecida Ludio Braga are acknowledged for the LLC-MK2 cells and for the advices during the manuscript preparation. The work was supported by grants and fellowships from FAPESP (2018/18257-1, 2018/15549-1, 2020/04923-0 to GP, 2013/07937-8 to LL, VDM (2021/00399-7) and SNM (2017/04032-5), CNPq (DOQ and “Productivity fellowship” to GP), and CAPES to LRF and JMDS.

References

- [1] C.B. Anfinsen, Principles that govern the folding of protein chains, *Science (New York, N.Y.)* 181(4096) (1973) 223-30.
- [2] P.A. Lund, The roles of molecular chaperones in vivo, *Essays in biochemistry* 29 (1995) 113-23.
- [3] S.G. Burston, A.R. Clarke, Molecular chaperones: physical and mechanistic properties, *Essays in biochemistry* 29 (1995) 125-36.
- [4] J. Buchner, Supervising the fold: functional principles of molecular chaperones, *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* 10(1) (1996) 10-9.
- [5] C. Xu, B. Bailly-Maitre, J.C. Reed, Endoplasmic reticulum stress: cell life and death decisions, *The Journal of clinical investigation* 115(10) (2005) 2656-64.
- [6] J.H. Lin, P. Walter, T.S. Yen, Endoplasmic reticulum stress in disease pathogenesis, *Annual review of pathology* 3 (2008) 399-425.
- [7] P. Walter, D. Ron, The unfolded protein response: from stress pathway to homeostatic regulation, *Science (New York, N.Y.)* 334(6059) (2011) 1081-6.
- [8] C.M. Dobson, The structural basis of protein folding and its links with human disease, *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 356(1406) (2001) 133-45.
- [9] P.J. Thomas, B.H. Qu, P.L. Pedersen, Defective protein folding as a basis of human disease, *Trends in biochemical sciences* 20(11) (1995) 456-9.
- [10] S.K. DebBurman, G.J. Raymond, B. Caughey, S. Lindquist, Chaperone-supervised conversion of prion protein to its protease-resistant form, *Proceedings of the National Academy of Sciences of the United States of America* 94(25) (1997) 13938-43.
- [11] M. Schröder, R.J. Kaufman, The mammalian unfolded protein response, *Annual review of biochemistry* 74 (2005) 739-89.
- [12] J. Hollien, Evolution of the unfolded protein response, *Biochimica et biophysica acta* 1833(11) (2013) 2458-63.
- [13] N. Hiramatsu, W.C. Chiang, T.D. Kurt, C.J. Sigurdson, J.H. Lin, Multiple Mechanisms of Unfolded Protein Response-Induced Cell Death, *The American journal of pathology* 185(7) (2015) 1800-8.
- [14] M. Jaud, C. Philippe, D. Di Bella, W. Tang, S. Pyronnet, H. Laurell, L. Mazzolini, K. Rouault-Pierre, C. Touriol, Translational Regulations in Response to Endoplasmic Reticulum Stress in Cancers, *Cells* 9(3) (2020).
- [15] J.D. Malhotra, R.J. Kaufman, The endoplasmic reticulum and the unfolded protein response, *Seminars in cell & developmental biology* 18(6) (2007) 716-31.
- [16] S.S. Vembar, J.L. Brodsky, One step at a time: endoplasmic reticulum-associated degradation, *Nature reviews. Molecular cell biology* 9(12) (2008) 944-57.
- [17] J. Hoseki, R. Ushioda, K. Nagata, Mechanism and components of endoplasmic reticulum-associated degradation, *Journal of biochemistry* 147(1) (2010) 19-25.
- [18] E.D. Werner, J.L. Brodsky, A.A. McCracken, Proteasome-dependent endoplasmic reticulum-associated protein degradation: an unconventional route to a familiar fate, *Proceedings of the National Academy of Sciences of the United States of America* 93(24) (1996) 13797-801.
- [19] L. Lemus, V. Goder, Regulation of Endoplasmic Reticulum-Associated Protein Degradation (ERAD) by Ubiquitin, *Cells* 3(3) (2014) 824-47.
- [20] W.S. Lee, W.H. Yoo, H.J. Chae, ER Stress and Autophagy, *Current molecular medicine* 15(8) (2015) 735-45.
- [21] C.M. Osowski, F. Urano, Measuring ER stress and the unfolded protein response using mammalian tissue culture system, *Methods in enzymology* 490 (2011) 71-92.
- [22] D.N. Itzhak, F. Sacco, N. Nagaraj, S. Tyanova, M. Mann, M. Murgia, SILAC-based quantitative proteomics using mass spectrometry quantifies endoplasmic reticulum stress in whole HeLa cells, *Disease models & mechanisms* 12(11) (2019).
- [23] R. Apweiler, H. Hermjakob, N. Sharon, On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database, *Biochimica et biophysica acta* 1473(1) (1999) 4-8.
- [24] A. Varki, R.D. Cummings, J.D. Esko, P. Stanley, G.W. Hart, M. Aebi, A.G. Darvill, T. Kinoshita, N.H. Packer, J.H. Prestegard, R.L. Schnaar, P.H. Seeberger, *Essentials of Glycobiology*, in: A. Varki, R.D. Cummings, J.D. Esko, P. Stanley, G.W. Hart, M. Aebi, A.G. Darvill, T. Kinoshita, N.H. Packer, J.H. Prestegard, R.L. Schnaar, P.H.

- Seeberger (Eds.), *Essentials of Glycobiology*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), 2015.
- [25] M. Thaysen-Andersen, N.H. Packer, Site-specific glycoproteomics confirms that protein structure dictates formation of N-glycan type, core fucosylation and branching, *Glycobiology* 22(11) (2012) 1440-52.
- [26] P.M. Rudd, R.A. Dwek, Glycosylation: heterogeneity and the 3D structure of proteins, *Critical reviews in biochemistry and molecular biology* 32(1) (1997) 1-100.
- [27] M. Thaysen-Andersen, N.H. Packer, B.L. Schulz, Maturing Glycoproteomics Technologies Provide Unique Structural Insights into the N-glycoproteome and Its Regulation in Health and Disease, *Molecular & cellular proteomics : MCP* 15(6) (2016) 1773-90.
- [28] L.Y. Lee, C.H. Lin, S. Fanayan, N.H. Packer, M. Thaysen-Andersen, Differential site accessibility mechanistically explains subcellular-specific N-glycosylation determinants, *Frontiers in immunology* 5 (2014) 404.
- [29] Y.F. Huang, K. Aoki, S. Akase, M. Ishihara, Y.S. Liu, G. Yang, Y. Kizuka, S. Mizumoto, M. Tiemeyer, X.D. Gao, K.F. Aoki-Kinoshita, M. Fujita, Global mapping of glycosylation pathways in human-derived cells, *Developmental cell* (2021).
- [30] A. Varki, Biological roles of glycans, *Glycobiology* 27(1) (2017) 3-49.
- [31] A. Helenius, M. Aebi, Roles of N-linked glycans in the endoplasmic reticulum, *Annual review of biochemistry* 73 (2004) 1019-49.
- [32] M.R. Wormald, R.A. Dwek, Glycoproteins: glycan presentation and protein-fold stability, *Structure (London, England : 1993)* 7(7) (1999) R155-60.
- [33] M. Aebi, R. Bernasconi, S. Clerc, M. Molinari, N-glycan structures: recognition and processing in the ER, *Trends in biochemical sciences* 35(2) (2010) 74-82.
- [34] J. Sun, K.E. Duffy, C.T. Ranjith-Kumar, J. Xiong, R.J. Lamb, J. Santos, H. Masarapu, M. Cunningham, A. Holzenburg, R.T. Sarisky, M.L. Mbow, C. Kao, Structural and functional analyses of the human Toll-like receptor 3. Role of glycosylation, *The Journal of biological chemistry* 281(16) (2006) 11144-51.
- [35] M.A. Schmidt, L.W. Riley, I. Benz, Sweet new world: glycoproteins in bacterial pathogens, *Trends in microbiology* 11(12) (2003) 554-61.
- [36] F.Y. Tan, C.M. Tang, R.M. Exley, Sugar coating: bacterial protein glycosylation and host-microbe interactions, *Trends in biochemical sciences* 40(7) (2015) 342-50.
- [37] J.W. Tams, J. Vind, K.G. Welinder, Adapting protein solubility by glycosylation. N-glycosylation mutants of *Coprinus cinereus* peroxidase in salt and organic solutions, *Biochimica et biophysica acta* 1432(2) (1999) 214-21.
- [38] J.A. Hanover, Glycan-dependent signaling: O-linked N-acetylglucosamine, *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* 15(11) (2001) 1865-76.
- [39] J. Tyson, L. Tranebjaerg, M. McEntagart, L.A. Larsen, M. Christiansen, M.L. Whiteford, J. Bathen, B. Aslaksen, S.J. Sørland, O. Lund, M.E. Pembrey, S. Malcolm, M. Bitner-Glindzicz, Mutational spectrum in the cardioauditory syndrome of Jervell and Lange-Nielsen, *Human genetics* 107(5) (2000) 499-503.
- [40] C. Reily, T.J. Stewart, M.B. Renfrow, J. Novak, Glycosylation in health and disease, *15(6)* (2019) 346-366.
- [41] T.T. Wang, IgG Fc Glycosylation in Human Immunity, *Current topics in microbiology and immunology* 423 (2019) 63-75.
- [42] S.P. Ferris, V.K. Kodali, R.J. Kaufman, Glycoprotein folding and quality-control mechanisms in protein-folding diseases, *Disease models & mechanisms* 7(3) (2014) 331-41.
- [43] P.M. Couto, J.J. Caramelo, Chapter 4 - Glycoprotein folding, in: V.B. Patel (Ed.), *Molecular Nutrition: Carbohydrates*, Academic Press 2019, pp. 59-71.
- [44] A. Ilari, C. Savino, Protein structure determination by x-ray crystallography, *Methods in molecular biology (Clifton, N.J.)* 452 (2008) 63-87.
- [45] K. Wüthrich, The way to NMR structures of proteins, *Nature structural biology* 8(11) (2001) 923-5.
- [46] Z.H. Zhou, Atomic resolution cryo electron microscopy of macromolecular complexes, *Advances in protein chemistry and structural biology* 82 (2011) 1-35.
- [47] A. Fontana, P. Polverino de Laureto, V. De Filippis, E. Scaramella, M. Zamboni, Probing the partly folded states of proteins by limited proteolysis, *Folding & design* 2(2) (1997) R17-26.
- [48] Y. Feng, G. De Franceschi, A. Kahraman, Global analysis of protein structural changes in complex proteomes, *32(10)* (2014) 1036-44.

- [49] M.D. Kazanov, Y. Igarashi, A.M. Eroshkin, P. Cieplak, B. Ratnikov, Y. Zhang, Z. Li, A. Godzik, A.L. Osterman, J.W. Smith, Structural determinants of limited proteolysis, *Journal of proteome research* 10(8) (2011) 3642-51.
- [50] S. Ninagawa, T. Okada, Y. Sumitomo, S. Horimoto, T. Sugimoto, T. Ishikawa, S. Takeda, T. Yamamoto, T. Suzuki, Y. Kamiya, K. Kato, K. Mori, Forcible destruction of severely misfolded mammalian glycoproteins by the non-glycoprotein ERAD pathway, *The Journal of cell biology* 211(4) (2015) 775-84.
- [51] S. Ninagawa, K. Mori, PNGase Sensitivity Assay to Study the Folding Status of Proteins, *Bio-protocol* 6(19) (2016) e1952.
- [52] R.N. Hull, W.R. Cherry, O.J. Tritch, Growth characteristics of monkey kidney cell strains LLC-MK1, LLC-MK2, and LLC-MK2(NCTC-3196) and their utility in virus research, *The Journal of experimental medicine* 115(5) (1962) 903-18.
- [53] R. Friedlander, E. Jarosch, J. Urban, C. Volkwein, T. Sommer, A regulatory link between ER-associated protein degradation and the unfolded-protein response, *Nature cell biology* 2(7) (2000) 379-84.
- [54] T. Yorimitsu, U. Nair, Z. Yang, D.J. Klionsky, Endoplasmic reticulum stress triggers autophagy, *The Journal of biological chemistry* 281(40) (2006) 30299-304.
- [55] X.Y. Xiang, X.C. Yang, J. Su, J.S. Kang, Y. Wu, Y.N. Xue, Y.T. Dong, L.K. Sun, Inhibition of autophagic flux by ROS promotes apoptosis during DTT-induced ER/oxidative stress in HeLa cells, *Oncology reports* 35(6) (2016) 3471-9.
- [56] U.K. Laemmli, Cleavage of structural proteins during the assembly of the head of bacteriophage T4, *Nature* 227(5259) (1970) 680-5.
- [57] A. Shevchenko, M. Wilm, O. Vorm, M. Mann, Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels, *Analytical chemistry* 68(5) (1996) 850-8.
- [58] A. Shevchenko, H. Tomas, J. Havlis, J.V. Olsen, M. Mann, In-gel digestion for mass spectrometric characterization of proteins and proteomes, *Nature protocols* 1(6) (2006) 2856-60.
- [59] M.L. Huynh, P. Russell, B. Walsh, Tryptic digestion of in-gel proteins for mass spectrometry analysis, *Methods in molecular biology* (Clifton, N.J.) 519 (2009) 507-13.
- [60] A.J. Alpert, Hydrophobic interaction chromatography of peptides as an alternative to reversed-phase chromatography, *Journal of chromatography* 444 (1988) 269-74.
- [61] P. Häggglund, J. Bunkenborg, F. Elortza, O.N. Jensen, P. Roepstorff, A new strategy for identification of N-glycosylated proteins and unambiguous assignment of their glycosylation sites using HILIC enrichment and partial deglycosylation, *Journal of proteome research* 3(3) (2004) 556-66.
- [62] S. Myslning, G. Palmisano, P. Højrup, M. Thaysen-Andersen, Utilizing ion-pairing hydrophilic interaction chromatography solid phase extraction for efficient glycopeptide enrichment in glycoproteomics, *Analytical chemistry* 82(13) (2010) 5598-609.
- [63] G. Palmisano, S.E. Lendal, K. Engholm-Keller, R. Leth-Larsen, B.L. Parker, M.R. Larsen, Selective enrichment of sialic acid-containing glycopeptides using titanium dioxide chromatography with analysis by HILIC and mass spectrometry, *Nature protocols* 5(12) (2010) 1974-82.
- [64] B.L. Parker, G. Palmisano, A.V. Edwards, M.Y. White, K. Engholm-Keller, A. Lee, N.E. Scott, D. Kolarich, B.D. Hambly, N.H. Packer, M.R. Larsen, S.J. Cordwell, Quantitative N-linked glycoproteomics of myocardial ischemia and reperfusion injury reveals early remodeling in the extracellular environment, *Molecular & cellular proteomics : MCP* 10(8) (2011) M110.006833.
- [65] M.J. Alves, R. Kawahara, R. Viner, W. Colli, E.C. Mattos, M. Thaysen-Andersen, M.R. Larsen, G. Palmisano, Comprehensive glycoproteomics of the epimastigote and trypomastigote stages of *Trypanosoma cruzi*, *Journal of proteomics* 151 (2017) 182-192.
- [66] J. Gobom, E. Nordhoff, E. Mirgorodskaya, R. Ekman, P. Roepstorff, Sample purification and preparation technique based on nano-scale reversed-phase columns for the sensitive analysis of complex peptide mixtures by matrix-assisted laser desorption/ionization mass spectrometry, *Journal of mass spectrometry : JMS* 34(2) (1999) 105-16.
- [67] J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification, *Nature biotechnology* 26(12) (2008) 1367-1372.
- [68] J. Cox, N. Neuhauser, A. Michalski, R.A. Scheltema, J.V. Olsen, M. Mann, Andromeda: a peptide search engine integrated into the MaxQuant environment, *Journal of proteome research* 10(4) (2011) 1794-805.

- [69] S. Tyanova, T. Temu, P. Sinitcyn, The Perseus computational platform for comprehensive analysis of (prote)omics data, *13*(9) (2016) 731-40.
- [70] R. Kornfeld, S. Kornfeld, Assembly of asparagine-linked oligosaccharides, *Annual review of biochemistry* 54 (1985) 631-64.
- [71] U. Raudvere, L. Kolberg, I. Kuzmin, T. Arak, P. Adler, H. Peterson, J. Vilo, g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update), *Nucleic acids research* 47(W1) (2019) W191-w198.
- [72] L.A. Kelley, S. Mezulis, C.M. Yates, M.N. Wass, M.J. Sternberg, The Phyre2 web portal for protein modeling, prediction and analysis, *10*(6) (2015) 845-58.
- [73] E. Bramucci, A. Paiardini, F. Bossa, S. Pascarella, PyMod: sequence similarity searches, multiple sequence-structure alignments, and homology modeling within PyMOL, *BMC bioinformatics* 13 Suppl 4(Suppl 4) (2012) S2.
- [74] G. Janson, C. Zhang, M.G. Prado, A. Paiardini, PyMod 2.0: improvements in protein sequence-structure analysis and homology modeling within PyMOL, *Bioinformatics (Oxford, England)* 33(3) (2017) 444-446.
- [75] G. Janson, A. Paiardini, PyMod 3: a complete suite for structural bioinformatics in PyMOL, *Bioinformatics (Oxford, England)* (2020).
- [76] R. Dass, F.A.A. Mulder, J.T. Nielsen, ODINPred: comprehensive prediction of protein order and disorder, *Scientific Reports* 10(1) (2020) 14780.
- [77] M.G. Yet, C.C. Chin, F. Wold, The covalent structure of individual N-linked glycopeptides from ovomucoid and asialofetuin, *The Journal of biological chemistry* 263(1) (1988) 111-7.
- [78] M. Thaysen-Andersen, S. Mysling, P. Højrup, Site-specific glycoprofiling of N-linked glycopeptides using MALDI-TOF MS: strong correlation between signal strength and glycoform quantities, *Analytical chemistry* 81(10) (2009) 3933-43.
- [79] S.Y. Chen, M. Dong, G. Yang, Glycans, Glycosite, and Intact Glycopeptide Analysis of N-Linked Glycoproteins Using Liquid Handling Systems, *92*(2) (2020) 1680-1686.
- [80] L.J. Huang, J.H. Lin, J.H. Tsai, Y.Y. Chu, Y.W. Chen, S.L. Chen, S.H. Chen, Identification of protein O-glycosylation site and corresponding glycans using liquid chromatography-tandem mass spectrometry via mapping accurate mass and retention time shift, *Journal of chromatography. A* 1371 (2014) 136-45.
- [81] M. Windwarder, F. Altmann, Site-specific analysis of the O-glycosylation of bovine fetuin by electron-transfer dissociation mass spectrometry, *Journal of proteomics* 108 (2014) 258-68.
- [82] V. Blanchard, M. Frank, B.R. Leeflang, R. Boelens, J.P. Kamerling, The structural basis of the difference in sensitivity for PNGase F in the de-N-glycosylation of the native bovine pancreatic ribonucleases B and BS, *Biochemistry* 47(11) (2008) 3435-46.
- [83] A.L. Tarentino, C.M. Gómez, T.H. Plummer, Jr., Deglycosylation of asparagine-linked glycans by peptide:N-glycosidase F, *Biochemistry* 24(17) (1985) 4665-71.
- [84] S. Hirani, R.J. Bernasconi, J.R. Rasmussen, Use of N-glycanase to release asparagine-linked oligosaccharides for structural analysis, *Analytical biochemistry* 162(2) (1987) 485-92.
- [85] I. Braakman, J. Helenius, A. Helenius, Manipulating disulfide bond formation and protein folding in the endoplasmic reticulum, *The EMBO journal* 11(5) (1992) 1717-22.
- [86] D.J. Opstelten, P. de Groot, M.C. Horzinek, H. Vennema, P.J. Rottier, Disulfide bonds in folding and transport of mouse hepatitis coronavirus glycoproteins, *Journal of virology* 67(12) (1993) 7394-401.
- [87] S. Allen, H.Y. Naim, N.J. Balleid, Intracellular Folding of Tissue-type Plasminogen Activator: EFFECTS OF DISULFIDE BOND FORMATION ON N-LINKED GLYCOSYLATION AND SECRETION (1995), *Journal of Biological Chemistry* 270(9) (1995) 4797-4804.
- [88] N.A. Cherepanova, S.V. Venev, J.D. Leszyk, S.A. Shaffer, R. Gilmore, Quantitative glycoproteomics reveals new classes of STT3A- and STT3B-dependent N-glycosylation sites, *218*(8) (2019) 2782-2796.
- [89] Y. Hu, P. Shah, D.J. Clark, Reanalysis of Global Proteomic and Phosphoproteomic Data Identified a Large Number of Glycopeptides, *90*(13) (2018) 8065-8071.
- [90] D. Nikolov, C. Li, M. Lackmann, P. Jeffrey, J. Himanen, Crystal structure of the human ephrin-A5 ectodomain, *Protein science : a publication of the Protein Society* 16(5) (2007) 996-1000.
- [91] G.J. Forse, M.L. Uson, F. Nasertorabi, A. Kolatkar, I. Lamberto, E.B. Pasquale, P. Kuhn, Distinctive Structure of the EphA3/Ephrin-A5 Complex Reveals a Dual Mode of Eph Receptor Interaction for Ephrin-A5, *PloS one* 10(5) (2015) e0127081.

- [92] J.-P. Himanen, M.J. Chumley, M. Lackmann, C. Li, W.A. Barton, P.D. Jeffrey, C. Vearing, D. Geleick, D.A. Feldheim, A.W. Boyd, M. Henkemeyer, D.B. Nikolov, Repelling class discrimination: ephrin-A5 binds to and activates EphB2 receptor signaling, *Nature Neuroscience* 7(5) (2004) 501-509.
- [93] S. Ferluga, R. Hantgan, Y. Goldgur, J.P. Himanen, D.B. Nikolov, W. Debinski, Biological and structural characterization of glycosylation on ephrin-A1, a preferred ligand for EphA2 receptor tyrosine kinase, *The Journal of biological chemistry* 288(25) (2013) 18448-57.
- [94] S. Goutham, I. Kumari, D. Pally, A. Singh, S. Ghosh, Y. Akhter, R. Bhat, Mutually exclusive locales for N-linked glycans and disorder in human glycoproteins, *Scientific Reports* 10(1) (2020) 6040.
- [95] T. Kubota, T. Shiba, S. Sugioka, S. Furukawa, H. Sawaki, R. Kato, S. Wakatsuki, H. Narimatsu, Structural basis of carbohydrate transfer activity by human UDP-GalNAc: polypeptide alpha-N-acetylgalactosaminyltransferase (pp-GalNAc-T10), *Journal of molecular biology* 359(3) (2006) 708-27.
- [96] T.A. Fritz, J.H. Hurley, L.B. Trinh, J. Shiloach, L.A. Tabak, The beginnings of mucin biosynthesis: the crystal structure of UDP-GalNAc:polypeptide alpha-N-acetylgalactosaminyltransferase-T1, *Proceedings of the National Academy of Sciences of the United States of America* 101(43) (2004) 15307-12.
- [97] C.L. Mendelsohn, E. Wimmer, V.R. Racaniello, Cellular receptor for poliovirus: molecular cloning, nucleotide sequence, and expression of a new member of the immunoglobulin superfamily, *Cell* 56(5) (1989) 855-65.
- [98] S. Koike, H. Horie, I. Ise, A. Okitsu, M. Yoshida, N. Iizuka, K. Takeuchi, T. Takegami, A. Nomoto, The poliovirus receptor protein is produced both as membrane-bound and secreted forms, *The EMBO journal* 9(10) (1990) 3217-24.
- [99] R. Lange, X. Peng, E. Wimmer, M. Lipp, G. Bernhardt, The poliovirus receptor CD155 mediates cell-to-matrix contacts by specifically binding to vitronectin, *Virology* 285(2) (2001) 218-27.
- [100] A. Halim, J. Nilsson, U. Rüetschi, C. Hesse, G. Larson, Human urinary glycoproteomics; attachment site specific analysis of N- and O-linked glycosylations by CID and ECD, *Molecular & cellular proteomics : MCP* 11(4) (2012) M111.013649.
- [101] A. Zibert, E. Wimmer, N glycosylation of the virus binding domain is not essential for function of the human poliovirus receptor, *Journal of virology* 66(12) (1992) 7368-73.
- [102] M. Udani, Q. Zen, M. Cottman, N. Leonard, S. Jefferson, C. Daymont, G. Truskey, M.J. Telen, Basal cell adhesion molecule/lutheran protein. The receptor critical for sickle cell adhesion to laminin, *The Journal of clinical investigation* 101(11) (1998) 2550-8.
- [103] G.P. Singh, Association between intrinsic disorder and serine/threonine phosphorylation in *Mycobacterium tuberculosis*, *PeerJ* 3 (2015) e724.
- [104] H. Nishi, K. Hashimoto, A.R. Panchenko, Phosphorylation in protein-protein binding: effect on stability and function, *Structure (London, England : 1993)* 19(12) (2011) 1807-15.
- [105] I. Nishikawa, Y. Nakajima, M. Ito, S. Fukuchi, K. Homma, K. Nishikawa, Computational prediction of O-linked glycosylation sites that preferentially map on intrinsically disordered regions of extracellular proteins, *International journal of molecular sciences* 11(12) (2010) 4991-5008.
- [106] E.T. Prates, X. Guan, Y. Li, X. Wang, P.K. Chaffey, M.S. Skaf, The impact of O-glycan chemistry on the stability of intrinsically disordered proteins, *Chemical science* 9(15) (2018) 3710-3715.
- [107] A.J. Parodi, Role of N-oligosaccharide endoplasmic reticulum processing reactions in glycoprotein folding and degradation, *The Biochemical journal* 348 Pt 1(Pt 1) (2000) 1-13.
- [108] S. Dejgaard, J. Nicolay, M. Taheri, D.Y. Thomas, J.J. Bergeron, The ER glycoprotein quality control system, *Current issues in molecular biology* 6(1) (2004) 29-42.
- [109] D. Shental-Bechor, Y. Levy, Effect of glycosylation on protein folding: a close look at thermodynamic stabilization, *Proceedings of the National Academy of Sciences of the United States of America* 105(24) (2008) 8256-61.
- [110] S.R. Hanson, E.K. Culyba, T.L. Hsu, C.H. Wong, J.W. Kelly, E.T. Powers, The core trisaccharide of an N-linked glycoprotein intrinsically accelerates folding and enhances stability, *Proceedings of the National Academy of Sciences of the United States of America* 106(9) (2009) 3131-6.
- [111] J.G. Beeley, Peptide chain conformation and the glycosylation of glycoproteins, *Biochemical and biophysical research communications* 76(4) (1977) 1051-5.
- [112] E. Bause, Structural requirements of N-glycosylation of proteins. Studies with proline peptides as conformational probes, *The Biochemical journal* 209(2) (1983) 331-6.

[113] P.V. Lam, R. Goldman, K. Karagiannis, T. Narsule, V. Simonyan, V. Soika, R. Mazumder, Structure-based comparative analysis and prediction of N-linked glycosylation sites in evolutionarily distant eukaryotes, *Genomics, proteomics & bioinformatics* 11(2) (2013) 96-104.