

TOOLS

# Quantitative glycoproteomics reveals new classes of STT3A- and STT3B-dependent N-glycosylation sites

Natalia A. Cherepanova<sup>1</sup>, Sergey V. Venev<sup>2</sup>, John D. Leszyk<sup>1,3</sup>, Scott A. Shaffer<sup>1,3</sup>, and Reid Gilmore<sup>1</sup> 

Human cells express two oligosaccharyltransferase complexes (STT3A and STT3B) with partially overlapping functions. The STT3A complex interacts directly with the protein translocation channel to mediate cotranslational glycosylation, while the STT3B complex can catalyze posttranslocational glycosylation. We used a quantitative glycoproteomics procedure to compare glycosylation of roughly 1,000 acceptor sites in wild type and mutant cells. Analysis of site occupancy data disclosed several new classes of STT3A-dependent acceptor sites including those with suboptimal flanking sequences and sites located within cysteine-rich protein domains. Acceptor sites located in short loops of multi-spanning membrane proteins represent a new class of STT3B-dependent site. Remarkably, the luminal ER chaperone GRP94 was hyperglycosylated in STT3A-deficient cells, bearing glycans on five silent sites in addition to the normal glycosylation site. GRP94 was also hyperglycosylated in wild-type cells treated with ER stress inducers including thapsigargin, dithiothreitol, and NGI-1.

## Introduction

Asparagine-linked glycosylation is one of the most common protein modification reactions in metazoan cells, occurring upon most proteins that enter the secretory pathway. The ER-localized oligosaccharyltransferase (OST) transfers a preassembled oligosaccharide onto asparagine residues in acceptor sites or sequons (N-X-T/S/C, X≠P) in nascent polypeptides. The genomes of metazoan organisms encode two STT3 proteins (STT3A and STT3B) that are the active site subunits of the two separate hetero-oligomeric OST complexes. The importance of both OST complexes for normal human health and development is highlighted by the diagnosis of patients with variants of congenital disorders of glycosylation (CDG; variants STT3A-CDG and STT3B-CDG) caused by reduced expression of STT3A or STT3B (Shrimal et al., 2013a).

The STT3A complex interacts directly with the protein translocation channel (Sec61 complex) to mediate cotranslational N-glycosylation of nascent polypeptides as they enter the lumen of the ER (Ruiz-Canada et al., 2009; Shrimal et al., 2017; Braunger et al., 2018). The STT3B complex is not associated with the translocon, yet glycosylates certain acceptor sites that have been skipped by the STT3A complex in either a cotranslational or posttranslocational mode (Ruiz-Canada et al., 2009; Shrimal et al., 2013b; Cherepanova et al., 2014). Acceptor sites in a folded protein domain cannot enter the OST active site (Lizak et al.,

2011; Wild et al., 2018). Glycosylation sites that are hypoglycosylated in STT3A-deficient cells are poor substrates for the STT3B complex (Ruiz-Canada et al., 2009), presumably due to limited access of the sequon to the STT3B active site. Previously identified STT3B-dependent sites include a subset of sequons with cysteine as the X-residue (Cherepanova et al., 2014) and sequons located in the C-terminal 50 residues of proteins (Shrimal et al., 2013b). MagT1, TUSC3, and their yeast orthologues (Ost3p and Ost6p) are oxidoreductases with an active site disulfide (CXXC motif) that are necessary for the full activity of the yeast OST (Schulz and Aebi, 2009; Schulz et al., 2009) and the mammalian STT3B complex (Cherepanova et al., 2014; Cherepanova and Gilmore, 2016). Current evidence suggests that the oxidoreductases recruit acceptor substrates and delay disulfide bond formation (Schulz et al., 2009; Cherepanova et al., 2014; Mohorko et al., 2014).

Our current understanding of the role of the STT3A and STT3B complexes in N-glycosylation is based primarily on pulse-chase labeling of glycoproteins in siRNA-treated HeLa cells (Ruiz-Canada et al., 2009; Shrimal and Gilmore, 2013; Shrimal et al., 2013b; Cherepanova et al., 2014). A limitation of this approach was that we were constrained to analyzing glycoproteins in a narrow size range (<80 kD) where loss of a single glycan could be detected by SDS-PAGE. Larger proteins with

<sup>1</sup>Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA; <sup>2</sup>Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, Worcester, MA; <sup>3</sup>Mass Spectrometry Facility, University of Massachusetts Medical School, Shrewsbury, MA.

Correspondence to Reid Gilmore: [reid.gilmore@umassmed.edu](mailto:reid.gilmore@umassmed.edu); N.A. Cherepanova's present address is Dept. of Psychiatry, University of Massachusetts Medical School, Shrewsbury, MA.

© 2019 Cherepanova et al. This article is distributed under the terms of an Attribution–Noncommercial–Share Alike–No Mirror Sites license for the first six months after the publication date (see <http://www.rupress.org/terms/>). After six months it is available under a Creative Commons License (Attribution–Noncommercial–Share Alike 4.0 International license, as described at <https://creativecommons.org/licenses/by-nc-sa/4.0/>).

more than six acceptor sites were not suitable for analysis due to poor resolution of protein glycoforms and the large number of mutants needed to define which sites were not glycosylated in siRNA-treated cells. Additional limitations of this strategy were the low number of proteins that could be tested in a single experiment and the incomplete depletion of STT3A or STT3B that was achieved using siRNAs.

Recently, the CRISPR/Cas9 gene-editing system was used to generate HEK293-derived knockout (KO) cell lines that lack the STT3A complex (*STT3A*<sup>-/-</sup>), the STT3B complex (*STT3B*<sup>-/-</sup>), or both oxidoreductases (*MAGT1/TUSC3*<sup>-/-</sup>; Cherepanova and Gilmore, 2016). Here, we use stable isotope labeling with amino acids (SILAC)-based glycoproteomics to quantify reductions in glycosylation site occupancy in these mutant cell lines. Quantitative acceptor site occupancy data were obtained for 900–1,100 sites in each mutant cell line. Loss of the STT3A complex reduces glycosylation of more sites than loss of the STT3B complex. Bioinformatics analysis identified new classes of STT3A- and STT3B-dependent acceptor sites. A small fraction of sites unexpectedly displayed increased site occupancy in the mutant cells. Five hyperglycosylated acceptor sites are located in the ER luminal chaperone GRP94. Pulse labeling experiments indicate that GRP94 is hyperglycosylated in cells that are exposed to OST inhibitors, low doses of ER stress-inducing agents, and are displayed by cells with partial or complete STT3A deficiencies. Our analysis provides insight into how the presence of two OST complexes enhances glycosylation site occupancy, thereby performing a major role in protein homeostasis in the ER.

## Results

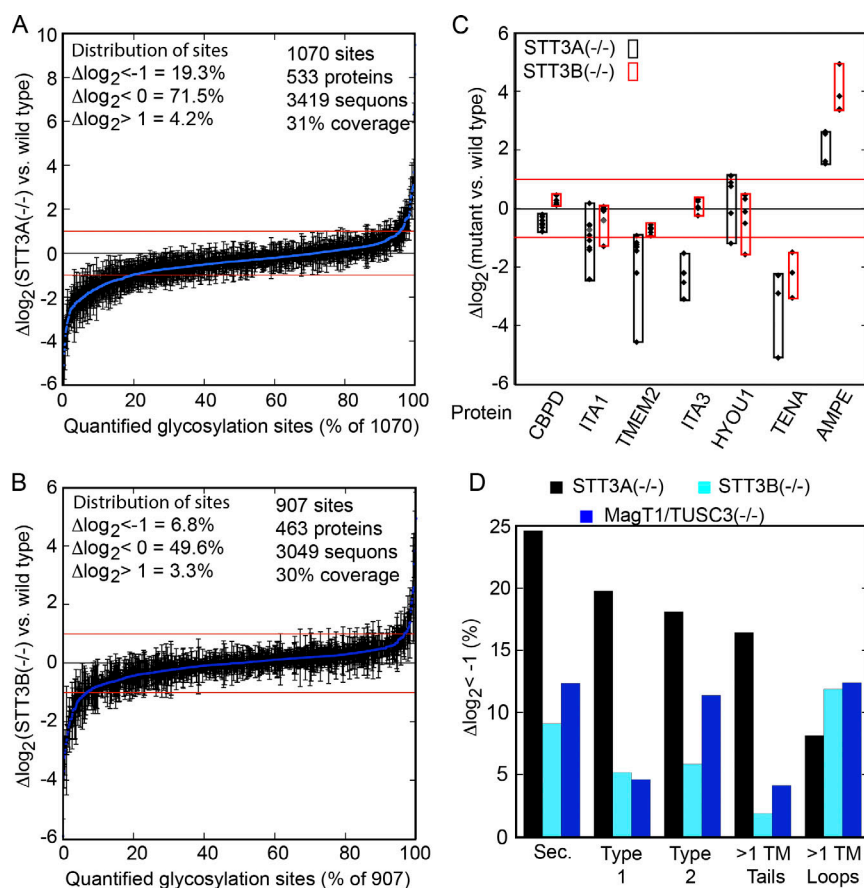
### Loss of the STT3A complex impacts more sequons than loss of the STT3B complex

WT or mutant HEK293 cells were cultured in the presence of media containing heavy amino acids (<sup>13</sup>C<sub>6</sub><sup>15</sup>N<sub>2</sub> Lys and <sup>13</sup>C<sub>6</sub> Arg) to uniformly label the proteins. Cell lysates from labeled WT and unlabeled mutant cells (or vice versa) were mixed at a 1:1 protein ratio, as determined using a protein assay, before trypsin digestion and subsequent glycopeptide isolation. The deglycoproteomics procedure we used was developed by Matthias Mann's laboratory (Zielinska et al., 2010). Tryptic glycopeptides are enriched with three lectins that recognize different terminal saccharides, followed by an N-glycanase digestion in the presence of H<sub>2</sub><sup>18</sup>O to remove N-linked glycans before liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS). The formerly glycosylated asparagine is detected by a 3D increase in peptide mass relative to the reference peptide sequence that occurs upon N-glycanase-dependent deamidation of asparagine in the presence of H<sub>2</sub><sup>18</sup>O. Peptides that do not contain a site that matches the consensus sequence (N-X-T/S/C, where X ≠ P) for N-glycosylation or are derived from proteins that lack an ER signal sequence or transmembrane (TM) span were excluded as false positives. The 2,190 bona fide glycosylation sites are derived from 892 proteins (Table S1) with a calculated coverage of 42.8% of the predicted glycosylation sites in the glycoproteins.

Quantitative glycosylation site occupancy data for *STT3A*<sup>-/-</sup> versus WT cells are presented as  $\Delta\log_2$  values (Fig. 1 A and Table S2) where a negative value reflects reduced site occupancy in the mutant cell line. Note that the  $\Delta\log_2$  value compares the site occupancy between cell lines, but does not provide an absolute measure of occupancy for the acceptor site in either cell line. Roughly 70% of the quantified sites showed reduced recovery in *STT3A* null cells ( $\Delta\log_2 < 0$ ), with 19% of acceptor sites showing more than a twofold decrease ( $\Delta\log_2 < -1$ ). A lower percentage of glycosylation sites showed reduced occupancy in cells that lack the STT3B complex (Fig. 1 B and Table S3) or lack both oxidoreductase subunits (*MAGT1/TUSC3*<sup>-/-</sup>; Fig. S1 A and Table S4). Less than 10% of the sites showed more than a twofold reduction in recovery in cells lacking a functional STT3B complex. Unexpectedly, 3–4% of the quantified glycopeptides showed twofold or greater increased recovery ( $\Delta\log_2 > 1$ ) in the mutant cell lines (Fig. 1, A and B; and Fig. S1 A). While some of the positive  $\Delta\log_2$  values are likely in error, most are based on multiple determinations that yielded similar results.

A reduction in the detection of a glycosite in the mutant cells could be explained by reduced glycosylation of that site or by reduced stable expression of the glycoprotein. To address this issue, we examined the  $\Delta\log_2$  scores for 84 proteins where we have quantitative data for three or more acceptor sites in both the STT3A and STT3B datasets (Fig. S2). Typical (CBPD, ITA1, TMEM2, and integrin  $\alpha$ -3) as well as atypical (HYOU1, tenascin A [TENA], and glutamyl aminopeptidase [AMPE]) examples are displayed in Fig. 1 C. For many proteins, the  $\Delta\log_2$  values are clustered near 0 (Fig. 1 C, CBPD; and Fig. S2). This pattern was more common for the STT3B dataset than for the STT3A dataset (Fig. S2). The STT3A  $\Delta\log_2$  values for acceptor sites in ITA1 and TMEM2 were spread over a broad range, consistent with the view that the quantified values primarily report on site occupancy rather than protein stability or expression. In contrast, the distribution of  $\Delta\log_2$  scores for integrin  $\alpha$ -3 suggests that this protein may have reduced stability in *STT3A*<sup>-/-</sup> cells. Three glycopeptides from HYOU1, a protein induced by the unfolded protein response (UPR) pathway, showed enhanced recovery in *STT3A*<sup>-/-</sup> cells, consistent with UPR induction (Ruiz-Canada et al., 2009; Cherepanova and Gilmore, 2016). TENA and AMPE displayed strongly reduced (TENA) or strongly enhanced (AMPE) recovery for all quantified sites in all three mutant cell lines (Fig. 1 C and Table S4).

Based on the National Center for Biotechnology Information (NCBI) annotations for N-terminal signal sequences and TM spans, we categorized the sites as being located in secretory/luminal proteins, type 1 (1 TM N<sub>lum</sub>-C<sub>cyt</sub>, i.e., one TM span and C terminus in the cytosol), type 2 (1 TM N<sub>cyt</sub>-C<sub>lum</sub>), or multi-TM membrane proteins (Tables S2 B, S3 B, and S4 B). Sites in multi-TM proteins were further categorized as being located in luminal loops or in N- or C-terminal tails. With the exception of sites in the loops of multi-TM proteins, loss of the STT3A complex caused a greater reduction in site occupancy (%  $\Delta\log_2 < -1$ ) than loss of the STT3B complex for all protein classes (Fig. 1 D). The percentage of sites impacted in *STT3B*<sup>-/-</sup> and *MAGT1/TUSC3*<sup>-/-</sup> cells was similar for most classes.



**Figure 1. Quantitative glycoproteomics of HEK293 derived cell lines.** Reduced N-glycosylation of acceptor sites in cell lines that lack STT3A (A) or STT3B (B). Differences in glycosylation site occupancy are expressed as  $\Delta\log_2$ . The blue lines are the mean  $\Delta\log_2$  values from two or more LC-MS/MS experiments. Error bars are SDs ( $n = 3-8$ ) or individual data points ( $n = 2$ ). The number of quantified sites, proteins from which they are derived, predicted sites, and calculated coverage are shown. (C) The  $\Delta\log_2$  values for selected glycoproteins with three or more quantified sites. Color-coded boxes indicate the dataset and highlight the  $\Delta\log_2$  range. (D) Percentage of glycosylation sites in different protein classes with  $\Delta\log_2 < -1$ .

The STT3A and STT3B datasets were merged to allow scatter plots for sites that were quantified in both cell lines (Fig. 2 A). The largest site category maps to the central sector, indicating that these sites can be glycosylated by STT3A or STT3B with no more than a twofold reduction in site occupancy. Glycosylation sites that are strongly STT3A-dependent (134 sites) or STT3B-dependent (39 sites) are the other major categories (Fig. 2 A). We obtained a different result when the STT3B and MAGT1/TUSC3 datasets were merged (Fig. 2 B). As expected (Fig. 1 B and Fig. S1 A), a higher percentage of sites was located in the central sector. Glycosylation sites that show reduced occupancy in both mutants (27 sites) are the second most abundant category. Compared with Fig. 2 A, the least squares line has a higher correlation coefficient ( $R$ ) and a steeper slope ( $S$ ). Loss of MagT1 and TUSC3 diminishes, but does not eliminate, the activity of the STT3B complex (Cherepanova et al., 2014), consistent with the slope of the least squares line ( $S < 1$ ).

Previously, we had shown that the STT3B complex was responsible for glycosylation of acceptor sites in the extreme C-terminal 50 residues of secretory proteins and type 2 membrane proteins (Shrimal et al., 2013b). Quantified glycosylation sites located in the C-terminal 150 residues of glycoproteins are shown in Fig. 2 C, with the symbol color designating the protein class, and the x axis specifying the distance in residues between the sequon and the C terminus. The quantitative glycoproteomics data confirmed and extended these findings by showing that extreme C-terminal sites are not impacted by loss of the STT3A complex but are hypoglycosylated in STT3B<sup>-/-</sup> and

MAGT1/TUSC3<sup>-/-</sup> cells. Glycosylation of extreme C-terminal sites in multi-spanning membrane proteins is not STT3B-dependent as these sites are either located in short C-terminal luminal tails or within the last luminal loop of the protein. These membrane-proximal sequons remain accessible to the translocation channel-associated STT3A complex following nascent chain termination.

### N-glycosylation of multi-TM proteins

Glycosylation sites in the STT3A dataset that are located in multi-TM proteins were categorized based on their location (Fig. 3 A and Table S2 B). The majority of strongly affected sites ( $\Delta\log_2 < -1$ ) are located in N-terminal tails (Fig. 3 A, inset). To obtain a better understanding of their STT3A dependence, the  $\Delta\log_2$  values for the N-terminal tail sites were plotted as a function of  $\log_{10}$  N-terminal tail length and color coded to indicate whether the site is an NXT site or an NXS site (Fig. 3 B). The observed trend is that longer N-terminal tail lengths correlate with STT3A dependence, with NXS sites being more sensitive to a STT3A deficiency than NXT sites.

Analysis of the multi-TM proteins in the STT3B dataset revealed that strongly affected sites ( $\Delta\log_2 < -1$ ) are located in luminal loops (Fig. 3 C). Since not all sites in loops were hypoglycosylated, we sought additional contributing factors (Fig. 3 D). The most severely affected sites were located in small loops of proteins with six or more TM spans. The affected sites clustered in the first luminal loop (inset in Fig. 3 D) and were enriched for NXS/C sites. The 10 STT3A-dependent sites that are

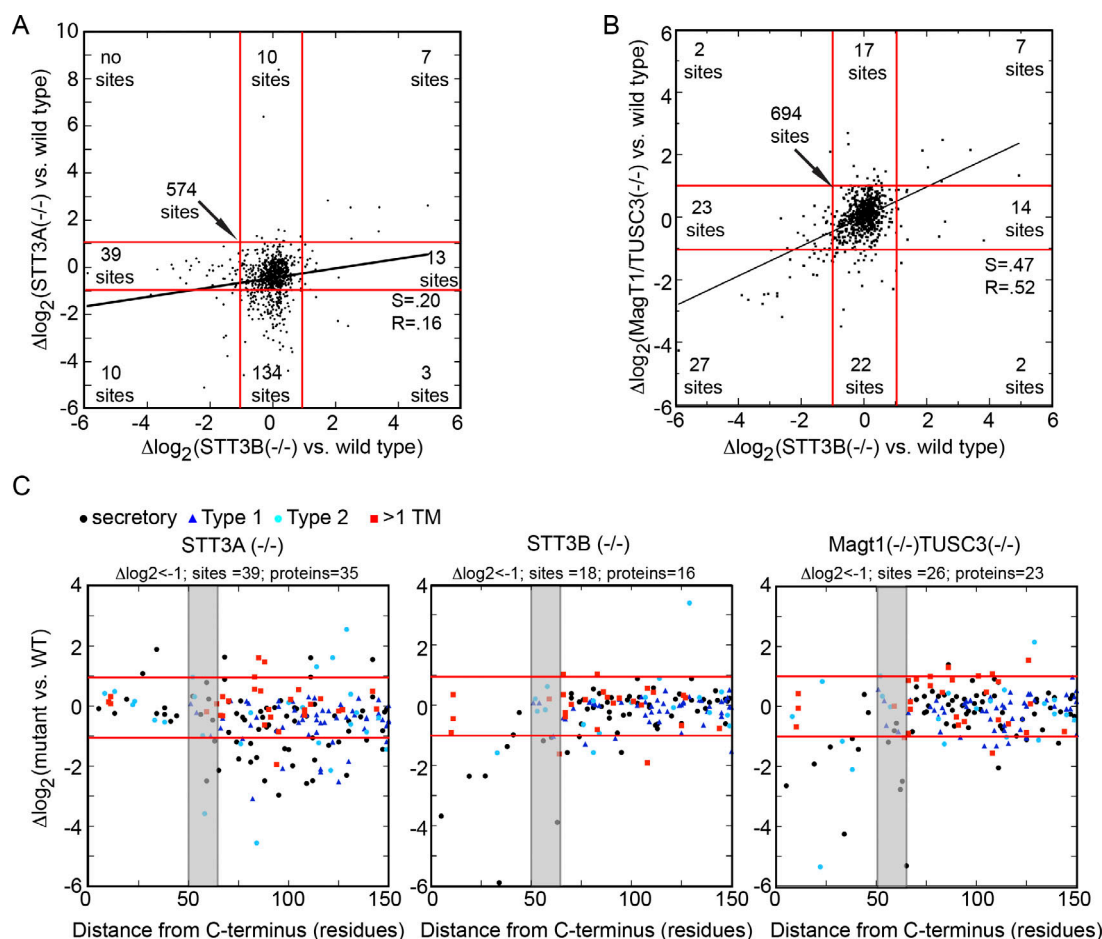


Figure 2. **Loss of the STT3A or STT3B impacts different sequons.** (A) Scatter plot of 790 sites shared by the *STT3A*<sup>-/-</sup> and *STT3B*<sup>-/-</sup> datasets. (B) Scatter plot of 808 sites shared by the *STT3B*<sup>-/-</sup> and *MAGT1/TUSC3*<sup>-/-</sup> datasets. (C) Glycosylation of sequons in the C-terminal 150 residues of proteins. Color codes designate protein classes. The number of C-terminal sites (C-terminal 150 residues,  $\Delta\log_2 < -1$ ) and proteins from which they are derived are shown above each panel. The gray shaded area designates a region (50–65 residues) of decreasing STT3B dependence (Shrimal et al., 2013b).

located in loops of multi-spanning membrane proteins (Fig. 3 A) are in larger loops than STT3B-dependent sites with no enrichment for six or more TM spans.

### Glycosylation of suboptimal sites

Assays using peptide substrates have shown that the OST has a higher affinity for NXT sites than NXS or NXC sites (Bause, 1984). We ordered the acceptor sites with respect to the value of  $\Delta\log_2$  and determined the frequency of NXT sites in each decile of the two databases (Fig. 4 A). The hypoglycosylated sites in both datasets are enriched for NXS/NXC sites. Nonetheless, >30% of the acceptor sites in deciles 6–10 are NXS sites, suggesting that other residues near the glycosylated asparagine impact acceptor site quality.

The X-residue side chain impacts the efficiency of glycosylation at least for NXS sites (Kasturi et al., 1995; Shakin-Eshleman et al., 1996; Malaby and Kobertz, 2014). To estimate how the X-residue influences sequon quality, we calculated an observed/expected (O/E) ratio at the +1 position for all NXT and NXS/C sites in our database of 2,190 experimentally verified HEK293 glycosylation sites (Fig. 4 B). This calculation is based on the hypothesis that sequences flanking a glycosylation site will

be selected to maximize acceptor site occupancy, particularly for sites that are important for protein stability or function. Expected residue distributions are derived from the bulk amino acid composition of the proteins in our dataset (Table S5). For both NXT sites and NXS sites, the composition of the X-residue deviates from the expected distribution, with NXS sites showing stronger negative selection against charged (E, D, R, and K), polar (Q and N), and bulky residues (W and M) at the X position. Although the rank order for NXT and NXS/C sites is not identical, the five residues showing positive selection (C, G, V, I, and Y) are the same for NXT and NXS/C sites, providing further evidence that the composition of the +1 residue is not random for both NXT and NXS/C sites.

Flanking residues that precede or follow the glycosylation site (-2, -1, and +3 positions) can contribute to glycosylation site quality due to interactions with the peptide-binding pocket of STT3A and STT3B. Bioinformatics analysis (Gavel and von Heijne, 1990) indicated that a proline residue at the +3 position is unfavorable, and this observation was confirmed by low (O/E) scores for proline at the +3 position of sequons (Table S5). Moreover, in vitro biosynthetic experiments indicate that the identity of the +3 residue impacts the glycosylation efficiency of



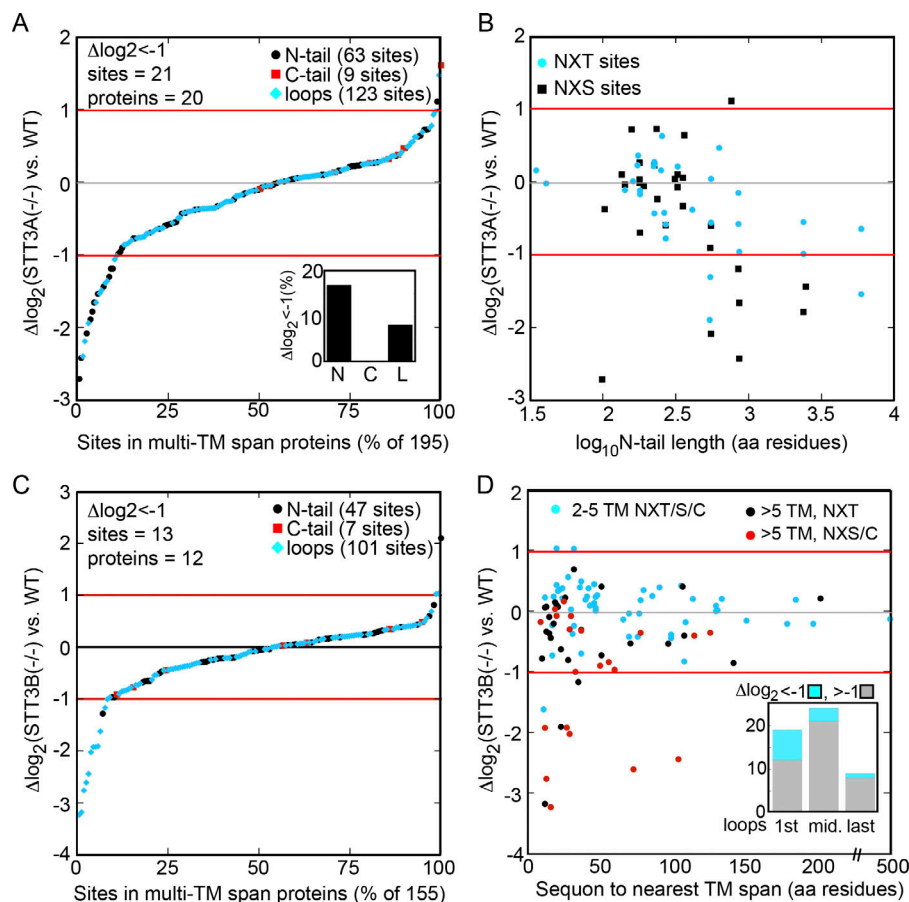


Figure 3. **Glycosylation of multi-TM proteins.**

(A) The  $\Delta\log_2$  values for sites located in loops (L) or N- or C-terminal tails (N- or C-tails, respectively) of multi-TM proteins in *STT3A*<sup>-/-</sup> cells. Sites in N-terminal tails are enriched in the negative zone (inset,  $\Delta\log_2 < -1$ ). (B) The  $\Delta\log_2$  values for sites in the N-terminal tails are plotted versus tail length ( $\log_{10}$ ) and color-coded for sequon type (NXT vs. NXS). (C) The  $\Delta\log_2$  values for sites located in loops or in N- or C-terminal tails of multi-TM proteins in *STT3B*<sup>-/-</sup> cells. (D) The  $\Delta\log_2$  values for sites located in loops are plotted versus the distance between the sequon and the closest TM span and color-coded for sequon type (NXT vs. NXS) and number of TM spans. Most of the affected sites ( $\Delta\log_2 < -1$ ) in proteins with more than five TM spans are located in the first luminal loop of the protein (see inset). For proteins with more than three luminal loops, the middle (mid.) category includes all luminal loops except the first and last. (A and C) The number of sites ( $\Delta\log_2 < -1$ ) and proteins from which they are derived is shown.

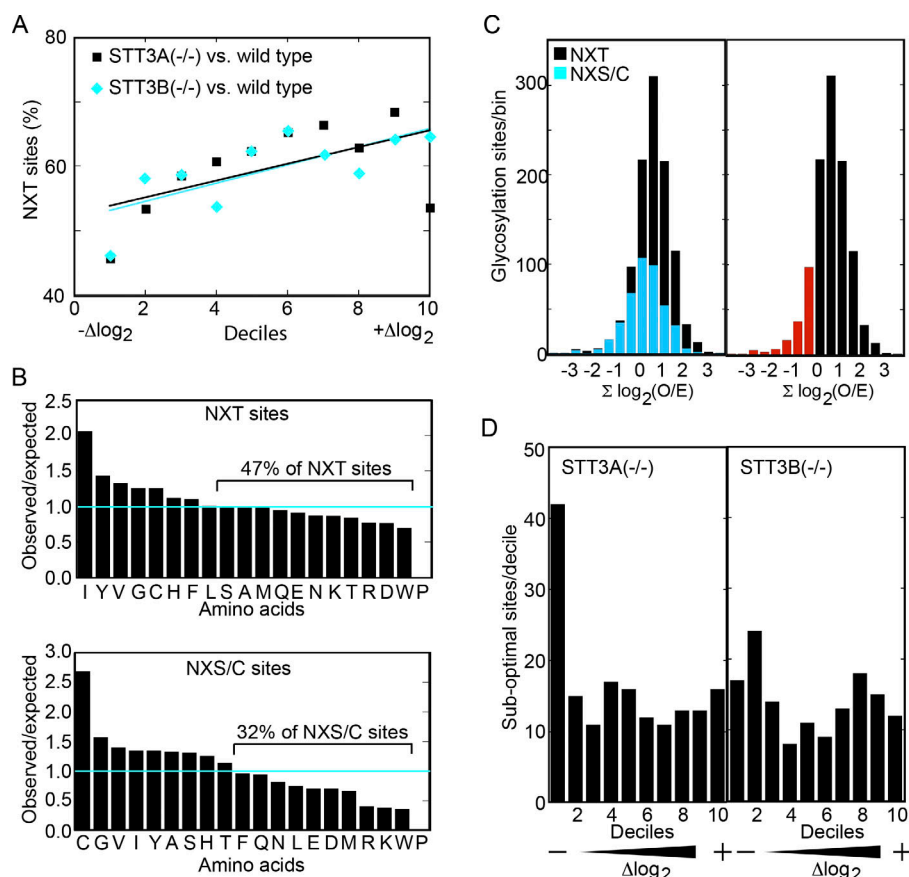
an NLS site (Mellquist et al., 1998). An interaction between the acidic -2 residue of a prokaryotic glycosylation site (D/E-Z-N-X-T/S, X≠P) and a basic residue in the peptide-binding pocket of a eubacterial OST (PglB) is essential for catalysis (Kowarik et al., 2006; Chen et al., 2007; Gerber et al., 2013). While eukaryotic OST sequons do not require an acidic residue at the -2 position, aromatic residues at the -2 position enhance N-glycosylation efficiency of a reporter protein in human cells (Murray et al., 2015). O/E ratios that were calculated for each residue in the minimal flanking sequence (-2, -1, +1, +2, and +3 residues) were converted into a flanking sequence score (Fig. 4 C,  $\Sigma \log_2$  [O/E]; and Tables S2 B and S3 B). Despite substantial overlap between NXT and NXS/C sites in the distribution of flanking sequence scores, most sequons with a flanking score below 0 are NXS/C sites (Fig. 4 C, left panel). Sequons with a negative flanking score (Fig. 4 C, right panel, red bars) are candidates for suboptimal sequons that may be poorly glycosylated. The marked enrichment of suboptimal sequons in the first decile of the *STT3A* dataset supports the hypothesis that cotranslational glycosylation by *STT3A* enhances glycosylation of suboptimal sequons (Fig. 4 D). The broader distribution of suboptimal sequons in the *STT3B* dataset suggests that the role for the *STT3B* complex may be limited to enhancing modification of suboptimal sites that are hypoglycosylated by *STT3A*. For example, both *STT3A* and *STT3B* contribute to the glycosylation of a poorly modified NAC site in CD69 (Shrimal et al., 2013b) that has a calculated flanking score of -3.3.

### N-glycosylation of cysteine-rich proteins

The *STT3A*-dependent substrates prosaposin and granulin are cysteine-rich glycoproteins that are composed of relatively small, cysteine-rich domains (Ruiz-Canada et al., 2009; Shrimal et al., 2013a). Since the average cysteine content of proteins in our database was 2.8%, we defined a cysteine-rich glycoprotein as having >4% cysteine. Using the 4% cutoff, 192 of the 892 proteins (~17%) in our dataset are defined as cysteine-rich glycoproteins. Roughly 17% of the sites in the *STT3A* and *STT3B* datasets are located in the cysteine-rich glycoproteins.

The majority of the quantified glycosylation sites (131 of 184) in the cysteine-rich glycoproteins have  $\Delta\log_2$  values below 0 in *STT3A*-deficient cells, suggesting that a high cysteine content correlates with *STT3A* dependence (Fig. 5 A). The glycosylation sites were further categorized as being located either within or outside a cysteine-rich domain. Most of the acceptor sites within cysteine-rich domains had  $\Delta\log_2 < 0$  (Fig. 5 A, cyan bars). Acceptor sites from cysteine-rich glycoproteins, regardless of their location, had a distribution that was similar to all sites in the *STT3B* dataset (Fig. 5 B).

LRP1 and mannose 6-phosphate receptor (MPRI) are two cysteine-rich glycoproteins that have multiple quantified sites (Fig. 5, C-E). The acceptor sites in LRP1 are distributed between cysteine-rich domains (LDLRA repeats), cysteine-free domains (LDLRB repeats), and intervening segments (Fig. 5 D). The acceptor sites located in the cysteine-rich domains or at domain boundaries were *STT3A*-dependent (Fig. 5 D, red circles), while



**Figure 4. N-glycosylation of suboptimal sequons.** (A) The STT3A and STT3B datasets were divided into deciles based on  $\Delta\log_2$  values. The percentage of NXT sites in each decile is shown. (B) The distribution of amino acids at the X position of 1,263 NXT sites and 906 NXS sites plus 21 NXC sites was determined. The results are expressed as an O/E ratio. O/E ratios were calculated for the  $-2$ ,  $-1$ ,  $+1(X)$ ,  $+2$ , and  $+3$  positions (Table S5). (C) Acceptor site flanking scores (Tables S2 B and S3 B) that were calculated as the sum of the  $\log_2$  O/E for the five residues ( $-2$  to  $+3$ ) are shown for the STT3A dataset. The number of NXT sites and NXS/C sites in half-unit bins is plotted (left panel). Glycosylation sites with a flanking score  $<0$  (red bars) are potential suboptimal sites. (D) The distribution of suboptimal sites in each decile of the STT3A and STT3B datasets.

the remaining sites were STT3A-independent (Fig. 5 D, black or blue circles). With the exception of a single site in LRP1, none of the sites has  $\Delta\log_2 < -0.5$  in the STT3B dataset. The luminal domain of MPRI is composed of fifteen 150-residue CIMR repeats, each with four disulfides (Fig. 5 E). All of the acceptor sites in MPRI have a  $\Delta\log_2 < 0$  in the STT3A dataset (Fig. 5 C). The observation that most MPRI sites show weaker STT3A dependence than the LRP1 sites located in LDLRA domains suggests that the size and folding kinetics of cysteine-rich domains as well as the sequon quality determine the extent to which specific sites can be modified by STT3B.

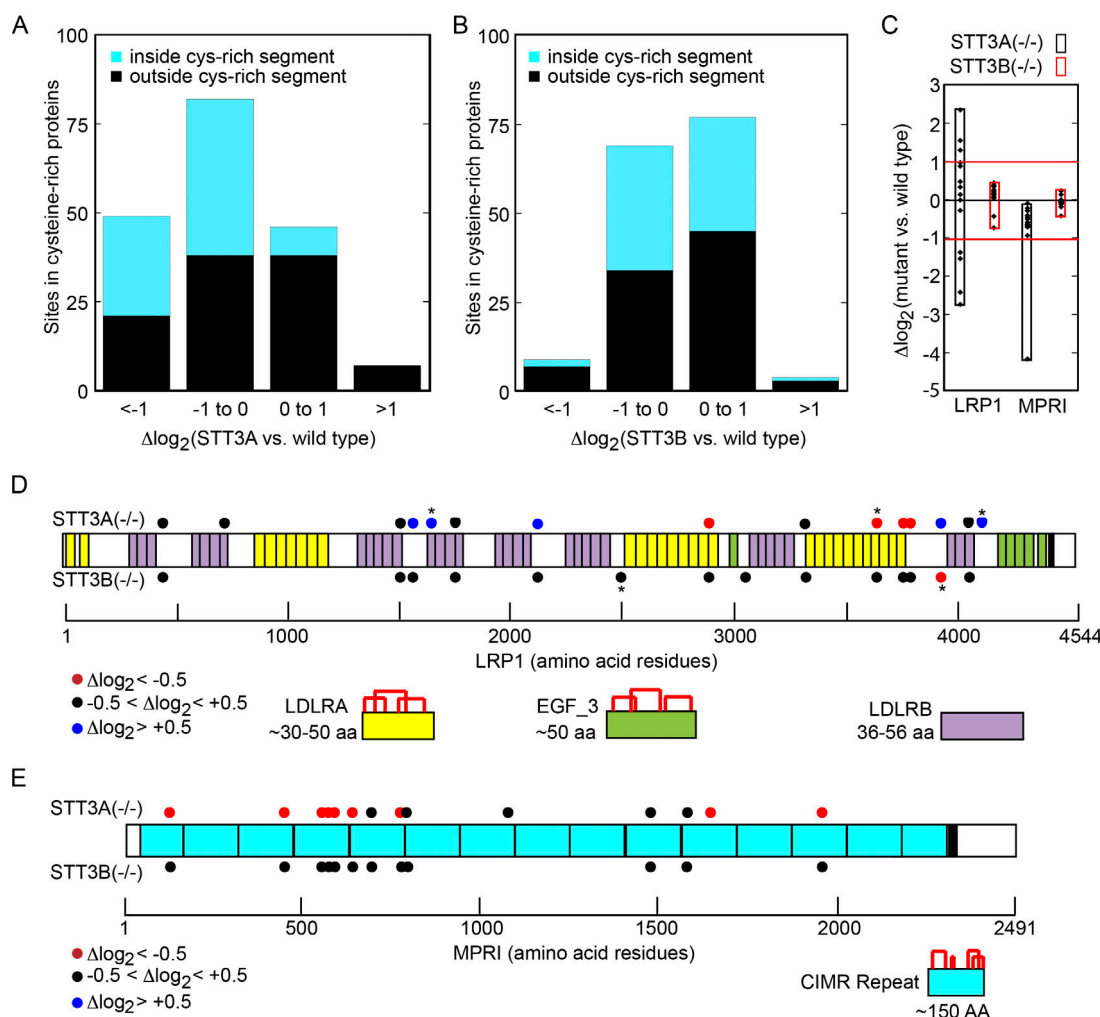
#### Pulse labeling of selected glycoproteins

Several glycoproteins were selected for pulse labeling with  $^{35}\text{S}$  methionine and cysteine to validate the glycoproteomics results. Color-coded symbols above each glycan summarize the SILAC results. Galectin 3 binding protein (LG3BP) has two extreme C-terminal sites and five internal sites (Fig. 6 A). Two LG3BP glycans are missing in STT3B $^{-/-}$  and MAGT1/TUSC3 $^{-/-}$  cells. As we lack coverage for three LG3BP sites in the STT3A dataset, we do not know which site is hypoglycosylated in the STT3A $^{-/-}$  cell line. The GLUT1 glucose transporter (GTR1) is a multi-spanning membrane protein that is hypoglycosylated in cells that lack STT3B. Pulse labeling confirmed reduced N-glycosylation of GLUT1 in STT3B-deficient cells (Fig. 6 B). Clusterin (CLUS1) was selected for pulse labeling because all quantified sites were strongly affected ( $\Delta\log_2 < -1$ ) in the STT3A $^{-/-}$  cells. Pulse labeling of clusterin in the STT3A $^{-/-}$  cells yielded glycoforms bearing two

to seven glycans. The lack of a functional STT3B complex did not reduce clusterin glycosylation. Despite having a cysteine content below 3%, the STT3A dependence of clusterin glycosylation might be related to formation of the five disulfides that link the N- and C-terminal halves of the protein.

MEGF9 was chosen as an example of a cysteine-rich glycoprotein with acceptor sites located in cysteine-rich and cysteine-free domains. MEGF9 is clearly hypoglycosylated in STT3A $^{-/-}$  cells (Fig. 6 D, left gel image). The MEGF9 5NQ mutant was constructed to examine glycosylation of the six sequons that are located in the EGF<sub>3</sub> repeats. When expressed in STT3A null cells, MEGF9 5NQ migrated near endoglycosidase H (EH)-digested MEGF9, indicating that most of the sites in the EGF<sub>3</sub> repeats are strongly STT3A-dependent (Fig. 6 D, right gel image). When disulfide bond formation is blocked by DTT treatment of cells, we observe an increased glycosylation of MEGF9 and MEGF9 5NQ in STT3A $^{-/-}$  cells, indicating that disulfide bond formation limits STT3B access to sequons in cysteine-rich domains.

The quantitative glycoproteomics results indicate that KDE2 has three STT3A-dependent sites including one site with a very negative flanking peptide score (KVN<sub>251</sub>GTP,  $\Sigma \log_2(\text{O/E}) = -2.39$ ) and a second site with a borderline score (KRN<sub>419</sub>LSD,  $\Sigma \log_2(\text{O/E}) = -0.007$ ). When synthesized in WT cells, KDE2DDK-His shows incomplete modification of one site (Fig. 6 E). Loss of the STT3A complex further reduces glycosylation, indicating that at least two sites are STT3A-dependent.



**Figure 5. Glycosylation of cysteine-rich proteins. (A and B)** Distribution of glycosylation sites in cysteine (cys)-rich domains of cysteine-rich proteins. Cysteine-rich domains are defined in Materials and methods. **(C)** The  $\Delta\log_2$  values for quantified sites in LRP1 and MPRI. Color-coded boxes specify the dataset and  $\Delta\log_2$  range. **(D)** Quantified acceptor sites in LRP1 are displayed as color-coded circles that flank the domain diagram showing cysteine-rich (LDLRA and EGF\_3) and cysteine-deficient (LDLRB) domains. Red lines designate disulfides. Glycosylation sites marked by asterisks were quantified in a single LC-MS/MS experiment. **(E)** Quantified acceptor sites in MPRI are shown as color-coded circles that flank the domain diagram showing the cysteine-rich CIMR repeats. TM spans (D and E) are indicated by black bars.

### Hyperglycosylation of GRP94

The predominant form of the ER luminal chaperone GRP94 (ENPL) has a single glycan attached to the N<sub>217</sub>DT site (Dersh et al., 2014) despite having six sequons (Fig. 7 A). Elimination of the STT3A complex caused a dramatic increase in glycosylation of the five silent sites, while loss of the STT3B complex had little impact on GRP94 glycopeptides (Fig. 7 B). A protein immunoblot experiment revealed that STT3A<sup>-/-</sup> cells lack the 1-glycan form of GRP94 but instead express a hyperglycosylated variant that is markedly less abundant than GRP94 in WT or STT3B-deficient cells (Fig. 7 C). STT3B<sup>-/-</sup> cells express zero and one glycan forms of GRP94.

Pulse-labeled GRP94 has three to five glycans in STT3A<sup>-/-</sup> cells as shown by an EH digestion time course (Fig. 7 D). GRP94 hyperglycosylation was replicated using an epitope-tagged expression construct, ruling out the possibility that the ENPL gene acquired an off-target mutation in the STT3A<sup>-/-</sup> cell line (Fig. 7 G). Elimination of the normal glycosylation site

(N217Q mutant) does not cause hyperglycosylation of GRP94 in WT cells, nor does it prevent GRP94 hyperglycosylation in STT3A<sup>-/-</sup> cells. When equal numbers of cells were used for pulse labeling (Fig. 7 E), the signal for GRP94 was strikingly elevated in STT3A<sup>-/-</sup> cells, consistent with induction of the UPR pathway (Ruiz-Canada et al., 2009; Cherepanova and Gilmore, 2016). The discrepancy between enhanced pulse labeling intensity (Fig. 7, D and E) and reduced steady-state expression (Fig. 7 C) indicates that hyperglycosylated GRP94 is degraded. Hyperglycosylated GRP94 synthesized by WT cells, albeit barely detectable (Fig. 7 E), is inactive, unfolded, and subject to degradation (Dersh et al., 2014). A pulse-chase labeling experiment showed that GRP94 hyperglycosylation occurred during the initial labeling pulse, followed by degradation that was apparent during the chase. Unglycosylated GRP94 synthesized by STT3B<sup>-/-</sup> cells is stable based on the comparison of protein immunoblot and pulse labeling results (Fig. 7, C and E).

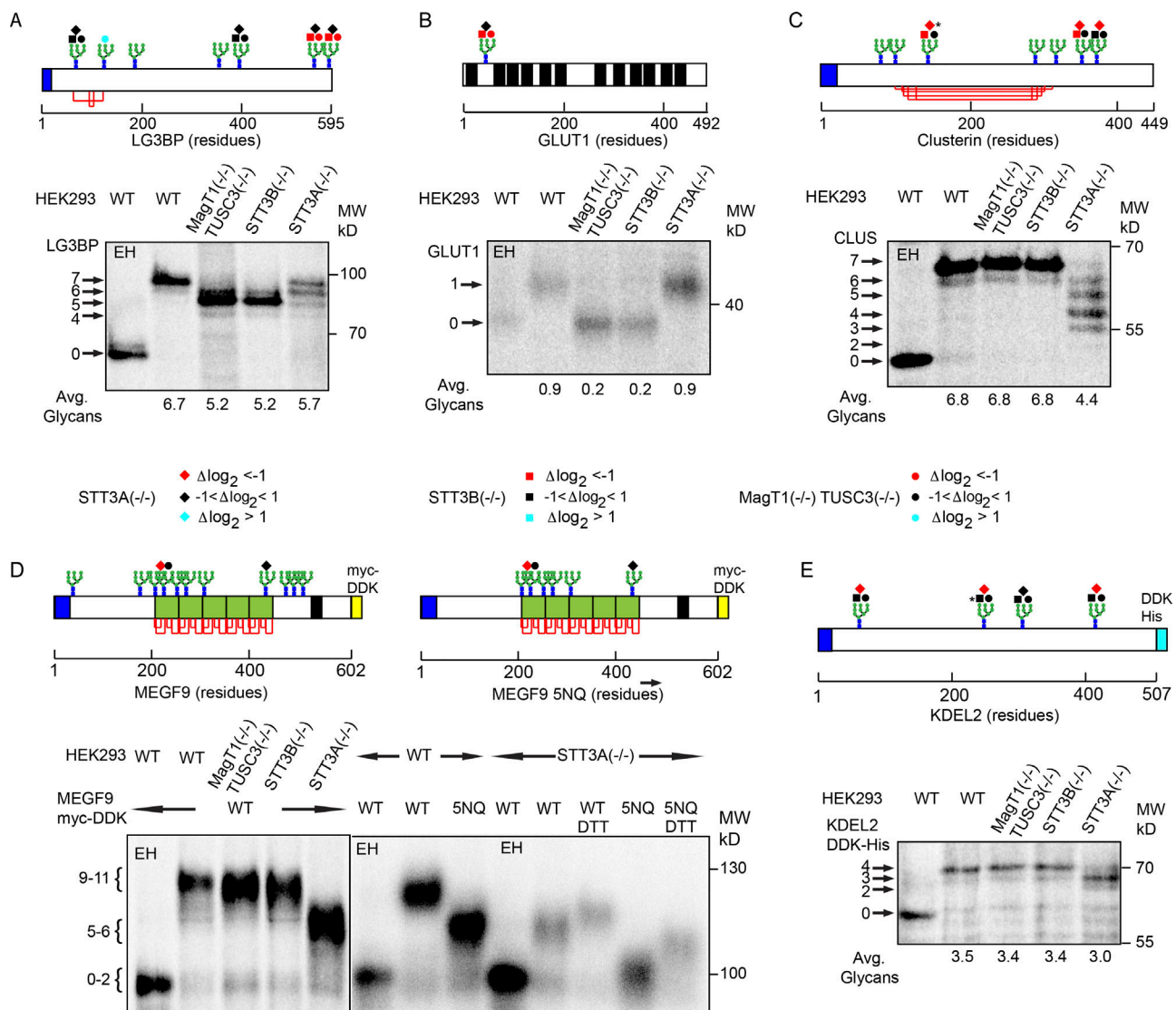


Figure 6. **Pulse labeling of selected glycoproteins. (A–E)** Blue and black segments designate signal sequences and TM spans respectively; red lines are disulfides. The disulfide pattern for KDEL2 is unknown. Color-coded symbols are defined below panels A–C. EH designates digestion with EH. Asterisks in panels C and E indicate sites that were quantified in a single LC-MS/MS experiment. Pulse labeled glycoproteins (10 min, <sup>35</sup>S Trans label) were immunoprecipitated, resolved by SDS-PAGE, and quantified by phosphor image analysis (see Materials and methods). **(A–C and E)** Quantification is for the experiment shown, which is representative of two or more pulse labeling experiments. **(D)** The MEGF9 5NQ mutant lacks the acceptor sites outside the EGF\_3 repeats (green blocks). As indicated, cells were preincubated with 3 mM DTT for 5 min before and during the pulse labeling period. The vertical line designates images derived from separate experiments aligned to show comparable mobilities for glycosylated MEGF9 and EH-digested MEGF9.

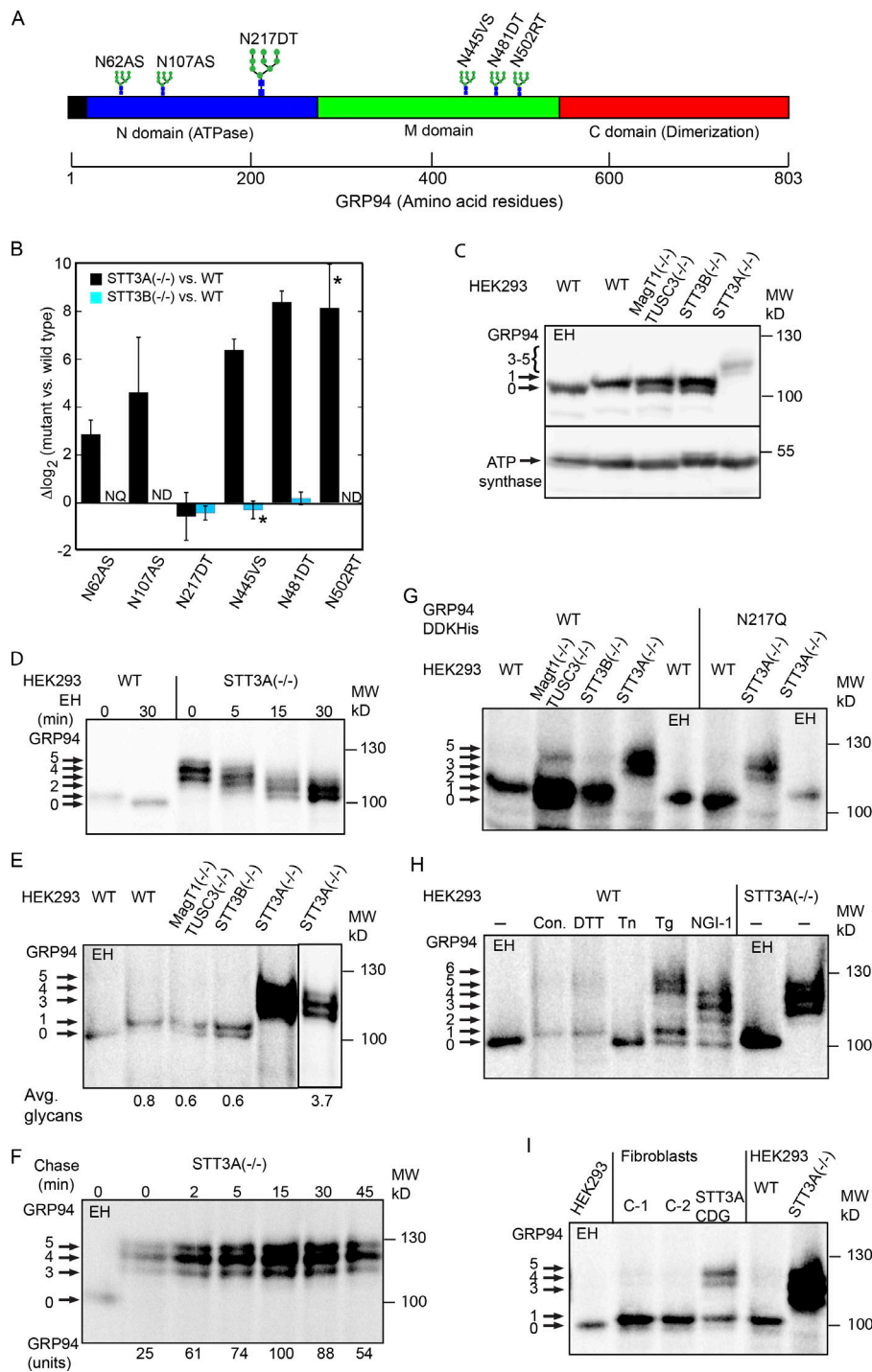
We next sought conditions that promote hyperglycosylation of GRP94 in WT cells. A 24-h treatment of cells with low concentrations of thapsigargin or DTT caused GRP94 hyperglycosylation (Fig. 7 H), albeit to a much lower extent with DTT. The ER-stress inducers did not eliminate the one-glycan form of GRP94. Treatment of cells with tunicamycin was uninformative, as the lipid-linked oligosaccharide pool was completely depleted as revealed by immunoprecipitation of prosaposin (data not shown). Treatment of cells with NGI-1, a recently identified OST inhibitor (Lopez-Sambrooks et al., 2016), yielded glycoforms having zero to five glycans. NGI-1 partially inhibits STT3A and STT3B, leading to a mild induction of the UPR pathway (Lopez-Sambrooks et al., 2016; Rinis et al., 2018). NGI-1 treatment of STT3B<sup>-/-</sup> cells also caused hyperglycosylation of GRP94 (Fig. S1 B).

The V626A point mutation in human STT3A causes a form of congenital disorders of glycosylation (STT3A-CDG) due to reduced stable expression of STT3A (Shrimal et al., 2013a; Ghosh et al., 2017). STT3A-CDG fibroblasts synthesize monoglycosylated active and hyperglycosylated inactive forms of GRP94 (Fig. 7 I).

## Discussion

The quantitative glycoproteomics analysis provided new insight into the role of the human OST complexes. Many of the quantified acceptor sites were located in proteins that were not suitable for pulse labeling analysis due to high molecular weight, numerous glycosylation sites, or a diffuse mobility when





analyzed by SDS-PAGE, so the quantitative glycoproteomics analysis filled a previous gap in our understanding of OST function. Due to the necessity of enriching the glycopeptides with lectins to maximize coverage of the glycoproteome, it was not feasible to systematically correct for the expression levels of the less abundant glycoproteins. For that reason, we want to stress that hypoglycosylation-dependent reductions in the stability of a specific glycoprotein may accentuate the apparent reduced site occupancy for acceptor sites in that glycoprotein. Pulse labeling analysis (Fig. 6) or targeted glycoproteomics to

simultaneously quantify site occupancy and protein expression levels could be used to analyze additional glycoproteins of interest.

#### In vivo roles of the STT3A and STT3B complexes

The protein translocation channel-associated STT3A complex (Shrimal et al., 2017; Braunger et al., 2018) glycosylates the majority of acceptor sites in human glycoproteins as indicated by negative  $\Delta\log_2$  values for 70% of the quantified sites in *STT3A*<sup>-/-</sup> cells. The predominant role for the STT3A complex in

N-glycosylation explains activation of the UPR pathway in STT3A- but not STT3B-deficient cells (Ruiz-Canada et al., 2009; Cherepanova and Gilmore, 2016). The three classes of STT3A-dependent acceptor sites that we identified in this study are (1) sequons with suboptimal flanking sequences, (2) sites in cysteine-rich protein domains, and (3) sites in long N-terminal tails of multi-TM proteins. Together these three categories account for roughly 50% of the acceptor sites that are strongly impacted ( $\Delta\log_2 < -1$ ) in STT3A<sup>-/-</sup> cells.

The three categories of STT3A-dependent glycosylation sites provide a logical framework for a general understanding of STT3A dependence. Based on the x-ray crystal structure of the eubacterial OST (PglB; Lizak et al., 2011) and the cryoelectron microscopy structures of eukaryotic OSTs (Bai et al., 2018; Braunger et al., 2018; Wild et al., 2018), the acceptor peptide binding site is located roughly 15 Å from the luminal membrane surface and can accommodate an acceptor site in an extended conformation. The ability of the STT3B complex to modify acceptor sites will therefore be limited by the folding kinetics of the region flanking the acceptor site, the affinity of the STT3B active site for the sequon, and the distance between the acceptor site and the luminal surface of the ER membrane. These factors likely contribute to the STT3A dependence of acceptor sites in large proteins including secretory proteins and type 1 and type 2 membrane proteins. Although we focused our analysis on small cysteine-rich domains like EGF<sub>3</sub> repeats and LDLRA domains since we can understand rapid formation of disulfide stabilized domains, we anticipate that a significant fraction of STT3A-dependent sites are located in noncysteine-rich protein domains that acquire conformations that are incompatible with sequon binding to the STT3B active site.

Fewer sequons were hypoglycosylated in cells that lack a functional STT3B complex, consistent with the conclusion that STT3B is primarily responsible for glycosylation of sites that have been skipped by the translocation channel associated STT3A complex. The two most prominent classes of STT3B-dependent sites are (1) extreme C-terminal sites in secretory proteins and type 2 membrane proteins and (2) acceptor sites located in small loops of multi-spanning membrane proteins. These two classes account for 40% of the strongly effected sites ( $\Delta\log_2 < -1$ ) in the STT3B dataset.

For glycosylation of multi-TM proteins, the luminal loop length and the distance between the sequon and the closest TM span are critical for two reasons. The peptide-binding site in STT3A is located 15 Å from the luminal surface of the ER membrane, and roughly 70 Å from the lateral gate of Sec61α (Braunger et al., 2018). Sequons in small luminal loops of multi-TM proteins like GLUT1 will have a low probability of engaging the STT3A active site before the entry of a subsequent TM span into the Sec61α transport pore mandates movement of the luminal loop away from the STT3A-Sec61 interface. Although the integration mechanism for multi-TM membrane proteins is not fully understood, we know that the TM spans of membrane proteins exit the lateral gate of Sec61α and move into the lipid bilayer in an N-terminal to C-terminal manner (Heinrich et al., 2000; Van den Berg et al., 2004). We conclude that acceptor sites that are located in larger luminal

loops of multi-TM proteins can access STT3A for cotranslational glycosylation.

Another objective of the glycoproteomics analysis was to determine whether a subset of acceptor sites that are glycosylated by the STT3B complex is insensitive to the absence of MagT1 and TUSC3. The N<sub>145</sub>QS and N<sub>354</sub>TS sites in clusterin were potential candidates for this type of site (Fig. 6 C). The results of the pulse labeling experiment for STT3B<sup>-/-</sup> cells did not confirm the glycoproteomics data. Inspection of the  $\Delta\log_2$  values for other sites that map into the same sector of the scatter plot (Fig. 2 B) did not disclose any other sites that were convincing candidates for STT3B-dependent, MagT1/TUSC3-independent sites in proteins that were suitable candidates for pulse labeling. Instead, the distribution of points in the scatter plot (Fig. 2 B) was consistent with the previous pulse labeling evidence that an oxidoreductase subunit is essential for full activity of the STT3B complex (Cherepanova et al., 2014; Cherepanova and Gilmore, 2016). A critical role for the MagT1/TUSC3 oxidoreductases in glycosylation of mammalian STT3B-dependent sites differs somewhat from the yeast OST complex, where the importance of the oxidoreductases (Ost3p or Ost6p) varies widely for glycosylation of specific sites (Schulz and Aebe, 2009; Schulz et al., 2009; Poljak et al., 2018).

### Suboptimal sequons and the STT3A complex

Despite the simple acceptor sequence required for N-linked glycosylation (N-X-T/S/C), certain sequons (e.g., many NXC sites) are poorly glycosylated, if at all, while other sequons are uniformly glycosylated. Here, we have obtained evidence that cells lacking either OST complex have a reduced ability to modify a subset of NXS/C sites. While previous studies have experimentally addressed the impact of the -2, +1, and +3 residues on the glycosylation of a specific sequon (Shakin-Eshleman et al., 1996; Mellquist et al., 1998; Malaby and Kobertz, 2014; Murray et al., 2015), it would not be feasible to test all possible flanking sequences (-2 to +3) for a single asparagine residue by pulse labeling. Based on the positive or negative selection of amino acids in flanking sequence residues, we derived a flanking sequence score. We have not incorporated weighting factors, so the flanking sequence score may overvalue the impact of the -2, -1, and +3 residues relative to the +1 and +2 residues. We conclude that cotranslational scanning of the nascent polypeptide by the STT3A complex enhances modification of suboptimal acceptor sites. Acceptor sites with suboptimal X residues are also poor substrates for the STT3B complex (Malaby and Kobertz, 2014), so a subset of sequons with negative flanking sequence scores is hypoglycosylated in the STT3A<sup>-/-</sup> cells.

### Hyperglycosylation of GRP94

The first point to consider is the functional impact of hyperglycosylation given that GRP94 is an essential protein required for normal development of mammalian tissues (as reviewed by Marzec et al., 2012). The GRP94 protein sequence is unusual in having multiple acceptor sites that are not appreciably glycosylated. The N107AS site is located in the ATP binding pocket of GRP94, with N107 directly contacting the adenine ring. The x-ray crystal structure of GRP94 (Dollins et al., 2007) indicates

low surface exposure of N502, so a glycan at this position will likely prevent folding of the M-domain. Glycosylation of either site will inactivate GRP94 and result in degradation, as reported previously for hyperglycosylated GRP94 (Dersh et al., 2014). Here, we obtained evidence that glycosylation of all five silent sites in GRP94 is dramatically elevated in *STT3A*<sup>-/-</sup> cells. Thus, despite UPR induction of GRP94 synthesis, GRP94 steady-state levels decrease in *STT3A*<sup>-/-</sup> cells relative to WT cells. We did not detect preferential degradation of specific glycoforms, so we lack evidence that *STT3A*<sup>-/-</sup> cells retain an active form of GRP94. Hyperglycosylation-induced degradation of GRP94 is likely detrimental to cells undergoing prolonged ER stress, as GRP94 is involved in the ER quality control and ER-associated degradation pathways (Christianson et al., 2008; Di et al., 2016).

The silent acceptor sites in GRP94 are conserved in mammalian GRP94 sequences (Fig. S3). Only two of the silent sites (N62 and N481) are missing from amphibian, fish, mollusk, and insect GRP94 sequences, suggesting that the silent sites serve a purpose. Cytosolic yeast and human HSP90 proteins have NXS sequences that align with the N<sub>107</sub>AS and N<sub>445</sub>VS sites in GRP94 (Fig. S3). N37 in *Saccharomyces cerevisiae* HSP90, the residue that aligns with N107 in GRP94, cannot be replaced with any other amino acid and retain biological activity (Mishra et al., 2016). A possible role for the silent sites in GRP94 would be to enhance glycan-dependent degradation of GRP94 variants that are aberrantly glycosylated at N107 or N445.

GRP94 was hyperglycosylated in *STT3A* null cells and *STT3A*-CDG cells or when WT or *STT3B*<sup>-/-</sup> cells are treated with NGI-1, an OST inhibitor that targets both *STT3A* and *STT3B* (Lopez-Sambrooks et al., 2016). Hyperglycosylation of GRP94 also correlated with conditions that cause protein-folding stress in the ER, including the 24-h treatments with low doses of DTT or thapsigargin. In contrast to the *STT3A*<sup>-/-</sup> cells, the drug treatments resulted in mixtures of nonglycosylated, monoglycosylated, and hyperglycosylated GRP94, so these cells retain active GRP94. Taken together, our current evidence suggests that hyperglycosylation of GRP94 is linked to ER protein folding stress. In contrast, cells that lack a fully functional *STT3B* complex hypoglycosylate GRP94.

What regulates GRP94 glycosylation to promote uniform skipping of the silent sites in nonstressed cells and efficient and rapid glycosylation of silent sites in stressed cells? None of the GRP94 sequons have a negative flanking sequence score, so these sites are not skipped because they are suboptimal. Instead, we speculate that there is a mechanism to restrict access of nascent GRP94 to the *STT3A* active site to block cotranslational glycosylation of the silent sites. RNC-Sec61 complexes that lack an adjacent OST complex have been detected by cryoelectron tomography of microsomes (Pfeffer et al., 2014; Braunger et al., 2018). Cotranslational glycosylation occurs by an N-terminal to C-terminal scanning mechanism (Chen et al., 1995; Shrimal and Gilmore, 2013) that is dependent on the nascent chain length (~65 residues) between the peptidyltransferase site on the ribosome and the acceptor binding site in *STT3A* (Whitley et al., 1996; Nilsson et al., 2003; Deprez et al., 2005). Consequently, the N<sub>62</sub>AS and N<sub>107</sub>AS sites in GRP94 could enter the *STT3A* active site before the normal glycosylation site (N<sub>217</sub>DT) is incorporated

into the nascent chain. For that reason, we can't invoke lectin recognition of the N217-glycan as a mechanism to block cotranslational glycosylation of the N<sub>107</sub>AS site. We suggest that the combination of UPR-induced GRP94 synthesis and chronic ER stress saturates a mechanism responsible for blocking glycosylation of the silent sites in GRP94 by *STT3A* and *STT3B*. While it is clear that the *STT3B* complex modifies the silent sites in *STT3A*<sup>-/-</sup> cells, GRP94 is also hyperglycosylated in *STT3B*<sup>-/-</sup> cells that have been treated with NGI-1. Thus, both OST complexes can efficiently glycosylate the silent GRP94 sites in stressed cells.

### STT3 gene duplication allows glycoproteome expansion

The ancient form of the eukaryotic OST complex is similar to the *STT3B* complex based on protein sequence alignments and expression of the *MagT1/TUSC3* orthologues in fungi and certain protists (Shrimal et al., 2013b). In addition to duplication of an ancestral *STT3* gene, interaction of an OST with the protein translocation channel required the generation of a gene encoding DC2, the subunit of the *STT3A* complex that directly contacts the Sec61 complex (Shrimal et al., 2017). DC2 occupies the same position in the mammalian *STT3A* complex that Ost3p occupies in the yeast OST complex (Braunger et al., 2018; Wild et al., 2018). Moreover, the three TM spans of DC2 are homologous to TM2-4 of Ost3p, and have a similar folded structure (Braunger et al., 2018; Wild et al., 2018).

Here, we have identified several classes of glycosylation sites that are primarily glycosylated by either the *STT3A* complex or the *STT3B* complex. The presence of two OST complexes in metazoan organisms enhances the efficiency of sequon modification, allowing glycosylation of diverse secretome proteins, and allows glycoproteome expansion to include acceptor sites in cysteine-rich proteins. Cooperation between the two OST complexes maximizes sequon occupancy in glycoproteins, which is essential for normal human health and development.

## Materials and methods

### Expression vectors

A full-length GRP94-verified clone was purchased from the Human Clone Collection at University of Massachusetts Medical School (ID 6165138). A KDEL2 clone was purchased from Transomics (ID 5266665). The coding sequences were amplified by PCR and cloned into the pCMV6-AC-DDK-His vector (Origene). A MEGF9-myc-DDK expression plasmid was purchased from Origene. To eliminate glycosylation sites, asparagine to glutamine mutations were introduced into the GRP94-DDK-His and MEGF9-myc-DDK expression vectors using standard site-directed mutagenesis to obtain the GRP94-DDK-His N217Q and MEGF9-myc-DDK 5NQ (N40Q, N182Q, N468Q, N481Q, and N500Q) mutants. All constructs were sequenced to confirm the desired mutations.

### Cell culture, transfection, and immunoblotting

The HEK293-derived *STT3A*<sup>-/-</sup>, *STT3B*<sup>-/-</sup>, and *MAGT1/TUSC3*<sup>-/-</sup> cell lines were characterized previously (Cherepanova and Gilmore, 2016). Briefly, the CRISPR/Cas9 system was used to



generate HEK293-derived cell lines that do not express STT3A, STT3B, or both MagT1 and TUSC3. The *STT3A*<sup>-/-</sup>, *STT3B*<sup>-/-</sup>, and *MAGT1/TUSC3*<sup>-/-</sup> KOs were genetically characterized by DNA sequencing of PCR products to identify the mutation sites in each of the targeted genes, which, with one exception, caused reading frame shifts accompanied by a nonsense codon. Protein immunoblotting using antisera specific for STT3A, STT3B, MagT1, and TUSC3 confirmed that the targeted proteins were not detectably expressed in the KO cell lines (Cherepanova and Gilmore, 2016).

HEK293 cells were cultured in 60-mm dishes at 37°C in DMEM (Gibco), 10% FBS with penicillin (100 U/ml), and streptomycin (100 µg/ml). Cells that were seeded at up to 80% confluency were transfected with reporter plasmids (6 µg) using Lipofectamine 2000 following a protocol from the manufacturer (Invitrogen) and were processed after 24 h. Primary skin fibroblasts were cultured in 100-mm plates and pulse labeled as described previously (Shrimal et al., 2013a).

### Antibodies and protein immunoblotting

Mouse monoclonal antibodies to GRP94 (MAB7606; R&D Systems), GLUT1 (ab40084; Abcam), the  $\alpha$ -subunit of the  $F_0F_1$ ATPase (612516; BD Biosciences), and the DDK epitope tag (F3165 anti-FLAG M2; Sigma-Aldrich) were obtained from commercial sources. Goat polyclonal antibodies for LG3BP and clusterin were obtained from R&D Systems (AF2226 and AF2937, respectively). Expression of GRP94 in cells was analyzed by protein immunoblotting as described previously using the  $\alpha$ -subunit of the  $F_0F_1$ -ATP synthase as the loading control (Cherepanova et al., 2014).

### Pulse-chase radiolabeling and immunoprecipitation

Cells expressing MEGF9-DDK-His or the MEGF9-DDK-His 5NQ mutant were untreated or treated with 3 mM DTT for 5 min before a 10-min pulse and 10-min chase labeling period. WT HEK293 cells were treated with DTT (200 µM), tunicamycin (0.6 µM), thapsigargin (0.1 µM), or NGI-1 (10 µM) for 24 h before pulse labeling and maintained in the drug during the labeling period.

Cells were pulse or pulse chase labeled, and glycoproteins were immunoprecipitated as described previously (Shrimal et al., 2013a,b). The pulse and chase intervals were as follows: (1) GRP94, 5-min pulse and chased as indicated; (2) LG3BP, GLUT1, clusterin, and MEGF9, 10-min pulse and 10-min chase. The glycoprotein substrates were immunoprecipitated, and the protein glycoforms were resolved by SDS-PAGE. As indicated, immunoprecipitated proteins were digested with EH (New England Biolabs). Dry gels were exposed to a phosphor screen (Fujifilm), scanned in Typhoon FLA 9000, and quantified using ImageQuant to determine the distribution of glycoforms and the average number of glycans per chain.

### SILAC labeling procedure

The SILAC labeling procedure was performed as described previously (Ong and Mann, 2006). Briefly, the WT and KO HEK293 cells were cultured in 60-mm plates in SILAC DMEM (A1443101; Life Technologies) supplemented with L-glutamine,

dialyzed FBS (88212; Thermo Fisher Scientific), penicillin, and streptomycin (15140; Life Technologies) in the presence of light L-lysine (L9037; Sigma-Aldrich) and light L-arginine (A6969; Sigma-Aldrich), or heavy L-lysine (<sup>13</sup>C<sup>6</sup><sup>5</sup>N<sub>2</sub>; CNLM-291-0.25; Cambridge Isotope Laboratories) and heavy L-arginine (<sup>13</sup>C<sub>6</sub>; CLM-2265-H-0.1; Cambridge Isotope Laboratories). Media concentrations for added lysine and arginine were 0.798 mM and 0.398 mM, respectively. After five passages, full incorporation of the heavy amino acids was achieved as verified by LC-MS/MS of whole cell-derived tryptic peptides.

### Protein digestion and glycopeptide enrichment

Glycopeptides were enriched for mass spectrometry as described previously (Wiśniewski et al., 2009; Zielinska et al., 2010) with minor modifications. Briefly, WT and mutant cells were lysed with SDS lysis buffer (4% SDS and 0.1 M DTT in 0.1 M Tris-Cl, pH 7.6). The cell lysates were diluted with 200 µl of buffer A (8 M urea in 0.1 M Tris-Cl, pH 8.5) and transferred to Microcon YM-30 filters (42409; EMD Millipore). The lysates were centrifuged at 14,000 rpm for 15 min, washed twice with buffer A, and then incubated for 20 min in the dark with 100 µl of buffer A containing 0.05 M iodoacetamide. Filters were centrifuged for 10 min and washed three times with buffer A. The filters were then washed three times with 100 µl of buffer B (40 mM NH<sub>4</sub>HCO<sub>3</sub>). Finally, 4 µg of trypsin (90057; Life Technologies) in 40 µl of buffer B were added to each filter. The filters were covered with parafilm and incubated overnight at 37°C. After filtration to separate the tryptic peptides from trypsin, the concentration of the peptides was determined using a bicinchoninic protein assay. Based on the results of the protein assay, fresh samples of the heavy and light lysates were mixed to obtain 0.2 mg of protein in a 1:1 heavy/light ratio. The mixed protein sample was processed for trypsin digestion as described above to obtain the heavy and light tryptic peptide mixture.

The tryptic peptide mixture was transferred to a fresh Microcon YM-30 filter. Filters were washed 2× with 40 µl of binding buffer (1 mM CaCl<sub>2</sub>, 1 mM MnCl<sub>2</sub>, 0.5 M NaCl, and 20 mM Tris-Cl, pH 7.5), and the combined fractions were adjusted to 1 mM PMSF and 1 mM tosyl-L-lysyl chloromethane hydrochloride (sc-201296; Santa Cruz Biotechnology). Samples were incubated for 10 min and divided into two aliquots. One aliquot was incubated with a lectin mixture containing 200 µg of concanavalin A (C2010; Sigma-Aldrich) and 200 µg of wheat germ agglutinin (L9640; Sigma-Aldrich) in 40 µl of 2× binding buffer. The second aliquot was incubated with 200 µg of ricin (RCA120; L7886; Sigma-Aldrich) in 40 µl of phosphate buffered saline. The combination of the three lectins will capture N-linked glycans with terminal  $\alpha$ -mannose,  $\alpha$ -glucose, N-acetylglucosamine, sialic acid, and galactose. After 1 h of incubation at room temperature, the samples were transferred to fresh YM-30 filters and centrifuged at 14,000 rpm for 10 min. The filters were washed 4× with 200 µl of binding buffer, and 2× with 40 µl of buffer C (40 mM NH<sub>4</sub>HCO<sub>3</sub> in H<sub>2</sub><sup>18</sup>O; 329878; Sigma-Aldrich). Peptides were deglycosylated by adding 2 µl of PNGase F (New England Biolabs) to each filter in 40 µl of buffer C and incubating for 3 h at 37°C. The deglycosylated peptides were eluted with two 50-µl washes of buffer B. The glycopeptide



mixtures were prepared and analyzed by LC-MS/MS four times (WT vs. *MAGT1/TUSC3*<sup>-/-</sup>), six times (WT vs. *STT3B*<sup>-/-</sup>), or eight times (WT vs. *STT3A*<sup>-/-</sup>) to obtain ~1,000 glycosylation sites that were quantified in two or more LC-MS/MS experiments. For each cell line comparison, the WT cells were cultured in heavy amino acids or light amino acids in different experiments to ensure that media differences do not cause apparent differences in glycosylation site occupancy.

Lectin enrichment of the glycopeptides and inclusion of H<sub>2</sub><sup>18</sup>O during N-glycanase deamidation of N-glycosylated asparagines are essential steps in this procedure. Otherwise the list of putative glycopeptides is overwhelmed with peptides derived from abundant cytosolic proteins that had undergone spontaneous deamidation of asparagine.

### LC-MS/MS analysis

Labeled tryptic peptide digests were submitted in 20 µl of 25 mM ammonium bicarbonate and acidified with 5 µl of 5% trifluoroacetic acid (total volume of 25 µl). Peptide digests were separated on a NanoAcquity (Waters Corp) UPLC. In brief, a 2.5-µl injection in 5% acetonitrile containing 0.1% formic acid (vol/vol) was loaded at 4.0 µl/min for 4.0 min onto a 100-µm inner-diameter fused-silica precolumn, packed with 2 cm of 5 µm (200 Å) Magic C18AQ (Bruker-Michrom) particles. Peptides were separated and eluted using a 75-µm inner diameter analytical column, packed with 25 cm of 3 µm (100 Å) Magic C18AQ particles, terminating with a gravity-pulled tip. A linear gradient was used from 5% solvent A (water and 0.1% [vol/vol] formic acid) to 35% solvent B (acetonitrile and 0.1% formic acid) in 90 min. Ions were generated by positive electrospray ionization via liquid junction and analyzed by a Q Exactive (Thermo Fisher Scientific) hybrid mass spectrometer. Mass spectra were acquired over the mass/charge (*m/z*) 300–1,750 range at 70,000 resolution (*m/z* 200) using a maximum ion fill time of 30 ms and an automatic gain control target of 10<sup>6</sup>. Data-dependent acquisition employed selecting the top 10 most abundant precursor ions for MS/MS by higher-energy collisional dissociation fragmentation using an isolation width of 1.6 D, collision energy of 27, resolution of 17,500, maximum ion fill time of 110 ms, and automatic gain control target of 10<sup>5</sup>. Dynamic exclusion was applied to maximize peptide identifications using the following parameters: duration (30 s), intensity threshold (9.1 × 10<sup>3</sup>), underfill ratio (1.0%), apex trigger (disabled), isotopes (excluded), peptide match (preferred), and charge exclusion (unassigned, 1+ and >8+).

### Analysis of mass spectrometry data

Raw data files were peak-processed with Proteome Discoverer (version 2.1; Thermo Fisher Scientific) before a database search with Mascot (version 2.6; Matrix Science) against the SwissProt Human database. Search parameters included tryptic specificity considering up to two missed cleavages. Variable modifications of oxidized methionine, pyroglutamic acid (N-terminal glutamine), N-terminal acetylation, deamidation of asparagine, deamidation of asparagine, and <sup>18</sup>O labeling combined, and fixed modifications for carbamidomethyl cysteine, were also considered. For SILAC experiments, incorporation of <sup>13</sup>C<sub>6</sub><sup>15</sup>N<sub>2</sub> lysine or

<sup>13</sup>C<sub>6</sub> arginine were allowed modifications. The mass tolerances were 10 ppm for the precursor and 0.05 D for the fragments. SILAC ratio quantitation was accomplished using Proteome Discoverer, and the results were loaded into Scaffold (version 4.8.9; Proteome Software) for peptide/protein identification and SILAC quantitation.

Peptides with identification probability scores exceeding 90% that contain N-deamidation (<sup>18</sup>O) were analyzed as potential glycopeptides. Complete protein sequences, signal sequence, and TM-span annotations were recovered from NCBI files for each peptide. Bona fide glycopeptides met the criteria of containing at least one canonical sequon (NXT/S/C, X≠P) and being located in a protein with an N-terminal signal sequence or one or more TM spans. Peptides with N-deamidation (<sup>18</sup>O) that lack canonical sequons were primarily derived from abundant cytosolic or nuclear proteins that are contaminants in the glycopeptide samples. Several glycopeptides that lack NCBI annotated signal sequences or TM spans were analyzed using the SignalP 4.1 server (Petersen et al., 2011) and the ΔG prediction server for TM-span identification (Hessa et al., 2007) and are included in the database if an ER targeting signal was evident. Mass spectra were inspected if peptides contained more sequons than deamidated (<sup>18</sup>O) asparagines or multiple asparagines to assign the correct acceptor site. All detected glycosylation sites in ER-targeted proteins are listed in Table S1. Glycopeptides that could be derived from more than one protein due to homology are listed a single time, with alternative protein assignments listed.

Spectral counts for the light and heavy versions of each glycopeptide were recovered from Scaffold files. Spectral counts for the N-glycanase digested glycopeptides were normalized based on the heavy/light ratio for nonglycopeptide contaminants in the analyzed sample, as this correction adjusts for any minor differences in the quantity of cell lysates that were mixed for the glycopeptide isolation. The Δlog<sub>2</sub> values were calculated for spectra and then averaged for those peptides where multiple quantified spectra were obtained in a single LC-MS/MS experiment. The SILAC datasets for WT vs. *STT3A*<sup>-/-</sup>, WT vs. *STT3B*<sup>-/-</sup>, and WT vs. *MAGT1/TUSC3*<sup>-/-</sup> are respectively tabulated in Tables S2, S3, and S4. If a glycosylation site was quantified in a single LC-MS/MS experiment, these sites are in Tables S2 D, S3 D, and S4 D and were excluded from further analysis. Glycosylation sites that were detected in two or more experiments were averaged, and SDs were calculated. Based on cutoff values for the SD of the Δlog<sub>2</sub> values, glycosylation sites were separated into a high confidence dataset (Tables S2 A, S3 A, and S4 A) or a low confidence dataset (Table S2 C, S3 C, and S4 C). If |Δlog<sub>2</sub>| < 1, sites with SD > 1 were assigned as low confidence. If |Δlog<sub>2</sub>| > 1, sites with SD > |Δlog<sub>2</sub>| were assigned as low confidence. Low confidence sites were excluded from further analysis.

### Bioinformatic analysis

The protein sequence, number and location of TM spans, and 20-amino acid sequence flanking the glycosylated asparagine were extracted from NCBI files for each quantified glycosylation site in the high-confidence datasets. When combined with the

location of the N terminus (cytosolic or luminal), this allowed membrane proteins to be classified as type 1 (1 TM N<sub>lum</sub>-C<sub>cyt</sub>), type 2 (1 TM N<sub>cyt</sub>-C<sub>lum</sub>), or multi-TM with a defined topology (e.g., 7 TM N<sub>lum</sub>-C<sub>cyt</sub>). Protein topology information is included in Tables S2 B, S3 B, and S4 B.

Cysteine-rich proteins were defined as having a cysteine content that exceeds 4%, based on the observation that no *S. cerevisiae* glycoprotein has a cysteine content exceeding 3.5%. A cysteine-rich segment in a cysteine-rich glycoprotein met at least one of the following criteria: (1) the site is located within a cysteine-rich domain recognized by the Prosite webserver (<https://prosite.expasy.org/>), (2) the 20-amino acid segment (from -9 to +10 relative to the glycosylated asparagine) contains three or more cysteine residues, or (3) the sequon has cysteine as the X-residue (NCT/S sites) or +2 residue (NXC).

The 2,190 glycosylation sites in our database (Table S1) are derived from 892 proteins. To estimate the total number of glycosylation sites in the 892 proteins, the protein sequences were scanned for N-X-T/S sites. N-X-C sites were not considered unless detected by LC-MS/MS, as the apparent modification frequency for NXC sites is low (Shrimal et al., 2013b). When acceptor site location is combined with protein topology analysis, we obtain the number of potential acceptor sites that are exposed to the ER lumen. Acceptor sites <10 residues from a predicted TM span were excluded as such sites are not glycosylated (Nilsson and von Heijne, 1993). The number of predicted glycosylation sites for each of the proteins is listed in Tables S1, S2 B, S3 B, and S4 B.

We used the 892 protein sequences (>710,000 residues) to estimate the bulk amino acid composition of the human secretome. The 20-residue sequences flanking the glycosylated asparagines were used to determine the observed composition of the flanking sequence residues relative to the expected (i.e., bulk) composition (Table S5). O/E ratios were calculated for the -2, -1, +1, +2, and +3 residues relative to the glycosylated asparagine, converted to log<sub>2</sub> values and summed to obtain a flanking sequence score. Cysteine content and flanking sequence scores are tabulated in Tables S2 B and S3 B.

### Online supplemental material

Fig. S1 A shows the distribution of  $\Delta\log_2$  values for the *MAGT1/TUSC3*<sup>-/-</sup> double mutant versus WT HEK293 cells. Fig. S1 B shows that GRP94 is hyperglycosylated in *STT3B*<sup>-/-</sup> cells that have been treated with NGI-1. Fig. S2 shows the distribution of  $\Delta\log_2$  scores for 75 glycoproteins with three or more quantified sites in both the STT3A and STT3B databases. Fig. S3 shows protein sequence alignments for metazoan GRP94 sequences in the vicinity of the glycosylation acceptor sites. Table S1 lists 2,190 glycosylation sites that were detected by LC-MS/MS during the course of this investigation. Table S2 is the SILAC-based glycoproteomics dataset for *STT3A*<sup>-/-</sup> versus WT HEK293 cells. Table S3 is the SILAC-based glycoproteomics dataset for *STT3B*<sup>-/-</sup> versus WT HEK293 cells. Table S4 is the SILAC-based glycoproteomics dataset for *MAGT1/TUSC3*<sup>-/-</sup> versus WT HEK293 cells. Table S5 includes the bulk amino acid composition of the 892 proteins in our dataset and the observed composition of the sequences that

flank NXT and NXS/C glycosylation sites. O/E ratios are shown for each of the flanking sequence residues.

## Acknowledgments

Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award No. GM43768.

The authors declare no competing financial interests.

Author contributions: R. Gilmore and N.A. Cherepanova conceived the study and designed experiments. N.A. Cherepanova performed experiments. J.D. Leszyk performed LC-MS/MS of N-glycanase-digested glycopeptides. S. V. Venev was involved in bioinformatic analysis. N. A. Cherepanova, S.A. Shaffer, and R. Gilmore wrote the manuscript.

Submitted: 1 April 2019

Revised: 13 June 2019

Accepted: 18 June 2019

## References

- Bai, L., T. Wang, G. Zhao, A. Kovach, and H. Li. 2018. The atomic structure of a eukaryotic oligosaccharyltransferase complex. *Nature*. 555:328–333. <https://doi.org/10.1038/nature25755>
- Bause, E. 1984. Model studies on N-glycosylation of proteins. *Biochem. Soc. Trans.* 12:514–517. <https://doi.org/10.1042/bst0120514>
- Braunger, K., S. Pfeffer, S. Shrimal, R. Gilmore, O. Berninghausen, E.C. Mandon, T. Becker, F. Förster, and R. Beckmann. 2018. Structural basis for coupling protein transport and N-glycosylation at the mammalian endoplasmic reticulum. *Science*. 360:215–219. <https://doi.org/10.1126/science.aar7899>
- Chen, M.M., K.J. Glover, and B. Imperiali. 2007. From peptide to protein: comparative analysis of the substrate specificity of N-linked glycosylation in *C. jejuni*. *Biochemistry*. 46:5579–5585. <https://doi.org/10.1021/bi602633n>
- Chen, W., J. Helenius, I. Braakman, and A. Helenius. 1995. Cotranslational folding and calnexin binding during glycoprotein synthesis. *Proc. Natl. Acad. Sci. USA*. 92:6229–6233. <https://doi.org/10.1073/pnas.92.14.6229>
- Cherepanova, N.A., and R. Gilmore. 2016. Mammalian cells lacking either the cotranslational or posttranslational oligosaccharyltransferase complex display substrate-dependent defects in asparagine linked glycosylation. *Sci. Rep.* 6:20946. <https://doi.org/10.1038/srep20946>
- Cherepanova, N.A., S. Shrimal, and R. Gilmore. 2014. Oxidoreductase activity is necessary for N-glycosylation of cysteine-proximal acceptor sites in glycoproteins. *J. Cell Biol.* 206:525–539. <https://doi.org/10.1083/jcb.201404083>
- Christianson, J.C., T.A. Shaler, R.E. Tyler, and R.R. Kopito. 2008. OS-9 and GRP94 deliver mutant alpha1-antitrypsin to the Hrd1-SEL1L ubiquitin ligase complex for ERAD. *Nat. Cell Biol.* 10:272–282. <https://doi.org/10.1038/ncb1689>
- Deprez, P., M. Gautschi, and A. Helenius. 2005. More than one glycan is needed for ER glucosidase II to allow entry of glycoproteins into the calnexin/calreticulin cycle. *Mol. Cell*. 19:183–195. <https://doi.org/10.1016/j.molcel.2005.05.029>
- Dersh, D., S.M. Jones, D. Eletto, J.C. Christianson, and Y. Argon. 2014. OS-9 facilitates turnover of nonnative GRP94 marked by hyperglycosylation. *Mol. Biol. Cell*. 25:2220–2234. <https://doi.org/10.1091/mbc.e14-03-0805>
- Di, X.J., Y.J. Wang, D.Y. Han, Y.L. Fu, A.S. Duerfeldt, B.S. Blagg, and T.W. Mu. 2016. Grp94 protein delivers gamma-aminobutyric acid type A (GABA) receptors to Hrd1 protein-mediated endoplasmic reticulum-associated degradation. *J. Biol. Chem.* 291:9526–9539. <https://doi.org/10.1074/jbc.M115.705004>
- Dollins, D.E., J.J. Warren, R.M. Immormino, and D.T. Gewirth. 2007. Structures of GRP94-nucleotide complexes reveal mechanistic differences between the hsp90 chaperones. *Mol. Cell*. 28:41–56. <https://doi.org/10.1016/j.molcel.2007.08.024>

- Gavel, Y., and G. von Heijne. 1990. Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites: implications for protein engineering. *Protein Eng.* 3:433–442. <https://doi.org/10.1093/protein/3.5.433>
- Gerber, S., C. Lizak, G. Michaud, M. Bucher, T. Darbre, M. Aebi, J.L. Reymond, and K.P. Locher. 2013. Mechanism of bacterial oligosaccharyltransferase: in vitro quantification of sequon binding and catalysis. *J. Biol. Chem.* 288:8849–8861. <https://doi.org/10.1074/jbc.M112.445940>
- Ghosh, A., J. Urquhart, S. Daly, A. Ferguson, D. Scotcher, A.A.M. Morris, and J. Clayton-Smith. 2017. Phenotypic heterogeneity in a congenital disorder of glycosylation caused by mutations in STT3A. *J. Child Neurol.* 32: 560–565. <https://doi.org/10.1177/0883073817696816>
- Heinrich, S.U., W. Mothes, J. Brunner, and T.A. Rapoport. 2000. The Sec61p complex mediates the integration of a membrane protein by allowing lipid partitioning of the transmembrane domain. *Cell.* 102:233–244. [https://doi.org/10.1016/S0092-8674\(00\)00028-3](https://doi.org/10.1016/S0092-8674(00)00028-3)
- Hessa, T., N.M. Meindl-Beinker, A. Bernsel, H. Kim, Y. Sato, M. Lerch-Bader, I. Nilsson, S.H. White, and G. von Heijne. 2007. Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature.* 450: 1026–1030. <https://doi.org/10.1038/nature06387>
- Kasturi, L., J.R. Eshleman, W.H. Wunner, and S.H. Shakin-Eshleman. 1995. The hydroxy amino acid in an Asn-X-Ser/Thr sequon can influence N-linked core glycosylation efficiency and the level of expression of a cell surface glycoprotein. *J. Biol. Chem.* 270:14756–14761. <https://doi.org/10.1074/jbc.270.24.14756>
- Kowarik, M., N.M. Young, S. Numao, B.L. Schulz, I. Hug, N. Callewaert, D.C. Mills, D.C. Watson, M. Hernandez, J.F. Kelly, et al. 2006. Definition of the bacterial N-glycosylation site consensus sequence. *EMBO J.* 25: 1957–1966. <https://doi.org/10.1038/sj.emboj.7601087>
- Lizak, C., S. Gerber, S. Numao, M. Aebi, and K.P. Locher. 2011. X-ray structure of a bacterial oligosaccharyltransferase. *Nature.* 474:350–355. <https://doi.org/10.1038/nature10151>
- Lopez-Sambrooks, C., S. Shrimal, C. Khodier, D.P. Flaherty, N. Rinis, J.C. Charest, N. Gao, P. Zhao, L. Wells, T.A. Lewis, et al. 2016. Oligosaccharyltransferase inhibition induces senescence in RTK-driven tumor cells. *Nat. Chem. Biol.* 12:1023–1030. <https://doi.org/10.1038/nchembio.2194>
- Malaby, H.L., and W.R. Kobertz. 2014. The middle X residue influences cotranslational N-glycosylation consensus site skipping. *Biochemistry.* 53: 4884–4893. <https://doi.org/10.1021/bi500681p>
- Marzec, M., D. Eletto, and Y. Argon. 2012. GRP94: An HSP90-like protein specialized for protein folding and quality control in the endoplasmic reticulum. *Biochim. Biophys. Acta.* 1823:774–787. <https://doi.org/10.1016/j.bbamer.2011.10.013>
- Mellquist, J.L., L. Kasturi, S.L. Spitalnik, and S.H. Shakin-Eshleman. 1998. The amino acid following an asn-X-Ser/Thr sequon is an important determinant of N-linked core glycosylation efficiency. *Biochemistry.* 37: 6833–6837. <https://doi.org/10.1021/bi972217k>
- Mishra, P., J.M. Flynn, T.N. Starr, and D.N.A. Bolon. 2016. Systematic mutant analyses elucidate general and client-specific aspects of Hsp90 function. *Cell Reports.* 15:588–598. <https://doi.org/10.1016/j.celrep.2016.03.046>
- Mohorko, E., R.L. Owen, G. Malojčić, M.S. Brozzo, M. Aebi, and R. Glockshuber. 2014. Structural basis of substrate specificity of human oligosaccharyl transferase subunit N33/Tusc3 and its role in regulating protein N-glycosylation. *Structure.* 22:590–601. <https://doi.org/10.1016/j.str.2014.02.013>
- Murray, A.N., W. Chen, A. Antonopoulos, S.R. Hanson, R.L. Wiseman, A. Dell, S.M. Haslam, D.L. Powers, E.T. Powers, and J.W. Kelly. 2015. Enhanced aromatic sequons increase oligosaccharyltransferase glycosylation efficiency and glycan homogeneity. *Chem. Biol.* 22:1052–1062. <https://doi.org/10.1016/j.chembiol.2015.06.017>
- Nilsson, I.M., and G. von Heijne. 1993. Determination of the distance between the oligosaccharyltransferase active site and the endoplasmic reticulum membrane. *J. Biol. Chem.* 268:5798–5801.
- Nilsson, I., D.J. Kelleher, Y. Miao, Y. Shao, G. Kreibich, R. Gilmore, G. von Heijne, and A.E. Johnson. 2003. Photocross-linking of nascent chains to the STT3 subunit of the oligosaccharyltransferase complex. *J. Cell Biol.* 161:715–725. <https://doi.org/10.1083/jcb.200301043>
- Ong, S.E., and M. Mann. 2006. A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nat. Protoc.* 1:2650–2660. <https://doi.org/10.1038/nprot.2006.427>
- Petersen, T.N., S. Brunak, G. von Heijne, and H. Nielsen. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods.* 8:785–786. <https://doi.org/10.1038/nmeth.1701>
- Pfeffer, S., J. Dudek, M. Gogala, S. Schorr, J. Linxweiler, S. Lang, T. Becker, R. Beckmann, R. Zimmermann, and F. Förster. 2014. Structure of the mammalian oligosaccharyl-transferase complex in the native ER protein translocon. *Nat. Commun.* 5:3072. <https://doi.org/10.1038/ncomms4072>
- Poljak, K., N. Selevsek, E. Ngwa, J. Grossmann, M.E. Losfeld, and M. Aebi. 2018. Quantitative profiling of N-linked glycosylation machinery in yeast *Saccharomyces cerevisiae*. *Mol. Cell. Proteomics.* 17:18–30. <https://doi.org/10.1074/mcp.RA117.000096>
- Rinis, N., J.E. Golden, C.D. Marceau, J.E. Carette, M.C. Van Zandt, R. Gilmore, and J.N. Contessa. 2018. Editing N-glycan site occupancy with small-molecule oligosaccharyltransferase inhibitors. *Cell Chem. Biol.* 25: 1231–1241.
- Ruiz-Canada, C., D.J. Kelleher, and R. Gilmore. 2009. Cotranslational and posttranslational N-glycosylation of polypeptides by distinct mammalian OST isoforms. *Cell.* 136:272–283. <https://doi.org/10.1016/j.cell.2008.11.047>
- Schulz, B.L., and M. Aebi. 2009. Analysis of glycosylation site occupancy reveals a role for Ost3p and Ost6p in site-specific N-glycosylation efficiency. *Mol. Cell. Proteomics.* 8:357–364. <https://doi.org/10.1074/mcp.M800219-MCP200>
- Schulz, B.L., C.U. Stirnimann, J.P. Grimshaw, M.S. Brozzo, F. Fritsch, E. Mohorko, G. Capitani, R. Glockshuber, M.G. Grütter, and M. Aebi. 2009. Oxidoreductase activity of oligosaccharyltransferase subunits Ost3p and Ost6p defines site-specific glycosylation efficiency. *Proc. Natl. Acad. Sci. USA.* 106:11061–11066. <https://doi.org/10.1073/pnas.0812515106>
- Shakin-Eshleman, S.H., S.L. Spitalnik, and L. Kasturi. 1996. The amino acid at the X position of an Asn-X-Ser sequon is an important determinant of N-linked core-glycosylation efficiency. *J. Biol. Chem.* 271:6363–6366. <https://doi.org/10.1074/jbc.271.11.6363>
- Shrimal, S., and R. Gilmore. 2013. Glycosylation of closely spaced acceptor sites in human glycoproteins. *J. Cell Sci.* 126:5513–5523. <https://doi.org/10.1242/jcs.139584>
- Shrimal, S., B.G. Ng, M.E. Losfeld, R. Gilmore, and H.H. Freeze. 2013a. Mutations in STT3A and STT3B cause two congenital disorders of glycosylation. *Hum. Mol. Genet.* 22:4638–4645. <https://doi.org/10.1093/hmg/ddt312>
- Shrimal, S., S.F. Trueman, and R. Gilmore. 2013b. Extreme C-terminal sites are posttranslocationally glycosylated by the STT3B isoform of the OST. *J. Cell Biol.* 201:81–95. <https://doi.org/10.1083/jcb.201301031>
- Shrimal, S., N.A. Cherepanova, and R. Gilmore. 2017. DC2 and KCP2 mediate the interaction between the oligosaccharyltransferase and the ER translocon. *J. Cell Biol.* 216:3625–3638. <https://doi.org/10.1083/jcb.201702159>
- Van den Berg, B., W.M. Clemons Jr., I. Collinson, Y. Modis, E. Hartmann, S.C. Harrison, and T.A. Rapoport. 2004. X-ray structure of a protein-conducting channel. *Nature.* 427:36–44. <https://doi.org/10.1038/nature02218>
- Whitley, P., I.M. Nilsson, and G. von Heijne. 1996. A nascent secretory protein may traverse the ribosome/endoplasmic reticulum translocase complex as an extended chain. *J. Biol. Chem.* 271:6241–6244. <https://doi.org/10.1074/jbc.271.11.6241>
- Wild, R., J. Kowal, J. Eyering, E.M. Ngwa, M. Aebi, and K.P. Locher. 2018. Structure of the yeast oligosaccharyltransferase complex gives insight into eukaryotic N-glycosylation. *Science.* 359:545–550. <https://doi.org/10.1126/science.aar5140>
- Wiśniewski, J.R., A. Zougman, N. Nagaraj, and M. Mann. 2009. Universal sample preparation method for proteome analysis. *Nat. Methods.* 6: 359–362. <https://doi.org/10.1038/nmeth.1322>
- Zielinska, D.F., F. Gnad, J.R. Wiśniewski, and M. Mann. 2010. Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. *Cell.* 141:897–907. <https://doi.org/10.1016/j.cell.2010.04.012>