Extreme C-terminal sites are posttranslocationally glycosylated by the STT3B isoform of the OST

Shiteshu Shrimal, Steven F. Trueman, and Reid Gilmore

Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01605

Metazoan organisms assemble two isoforms of the oligosaccharyltransferase (OST) that have different catalytic subunits (STT3A or STT3B) and partially nonoverlapping roles in asparagine-linked glycosylation. The STT3A isoform of the OST is primarily responsible for co-translational glycosylation of the nascent polypeptide as it enters the lumen of the endoplasmic reticulum. The C-terminal 65–75 residues of a glycoprotein will not contact the translocation channel–associated STT3A isoform of the OST complex before chain termination. Biosynthetic pulse labeling of five human glycoproteins showed that extreme C-terminal glycosylation sites were modified by an STT3B-dependent posttranslocational mechanism. The boundary for STT3B-dependent glycosylation of C-terminal sites was determined to fall between 50 and 55 residues from the C terminus of a protein. C-terminal NXT sites were glycosylated more rapidly and efficiently than C-terminal NXS sites. Bioinformatics analysis of glycopeptide databases from metazoan organisms revealed a lower density of C-terminal acceptor sites in glycoproteins because of reduced positive selection of NXT sites and negative selection of NXS sites.

Introduction

Asparagine-linked glycosylation is an evolutionarily conserved protein modification reaction that occurs on N-(X≠P)-T/S/C consensus sites (sequons) on newly synthesized proteins in the lumen of the ER of eukaryotic cells and on the exoplasmic surface of archaebacteria and certain proteobacteria (Larkin and Imperiali, 2011). The donor substrate for N-linked glycosylation in most eukaryotic organisms is the dolichol pyrophosphate–linked oligosaccharide $GlcNAc_2Man_9Glc_3$. The crystal structure of an acceptor peptide bound to the *Campylobacter lari* PglB, a eubacterial oligosaccharyltransferase (OST), showed that the hydroxyamino acid (T/S) in the sequon helps position the asparagine next to the active site residues involved in catalysis (Lizak et al., 2011). The architecture of the PglB active site is consistent with biochemical experiments, indicating that functional acceptor sites must be in unfolded or flexible regions of newly synthesized polypeptides (Kowarik et al., 2006).

In higher eukaryotes, the OST is a heteroligomeric membrane protein composed of seven to eight nonidentical subunits (Kelleher and Gilmore, 2006). The STT3 subunit contains the OST active site (Yan and Lennarz, 2002; Nilsson et al., 2003) and is the sole OST subunit that is conserved between the eukaryotic, archaebacterial, and eubacterial enzymes (Wacker et al., 2002). Vertebrate, insect, and multicellular plant genomes encode two STT3 proteins (STT3A and STT3B; Kelleher et al., 2003). The canine and human STT3A and STT3B proteins are incorporated into OST complexes that have distinct kinetic properties, with the STT3B isoform displaying a reduced stringency of selection of the dolichol pyrophosphate–linked $GlcNAc_2Man_9Glc_3$ donor relative to lumenally oriented assembly intermediates (dolichol pyrophosphate–linked $GlcNAc_2Man_{5-9}Glc_{0-2}$; Kelleher et al., 2003). Selective depletion of STT3A or STT3B in HeLa cells via siRNA treatment has shown that the two OST isoforms have partially nonoverlapping cellular roles in N-linked glycosylation (Ruiz-Canada et al., 2009). The STT3A isoform is associated with the protein translocation channel (Shibatani et al., 2005) and mediates co-translational glycosylation of NXT/S sites on a nascent polypeptide as it enters the ER lumen (Nilsson et al., 2003; Ruiz-Canada et al., 2009). Glycosylation of the rapidly folding protein prosaposin is particularly sensitive to STT3A depletion (Ruiz-Canada et al., 2009). The STT3B isoform of the OST complex can mediate posttranslational

The Rockefeller University Press $30.00
J. Cell Biol. Vol. 201 No. 1 81–95
www.jcb.org/cgi/doi/10.1083/jcb.201301031

JCB 81

glycosylation of NXT/S sites that are skipped by the translocation channel–associated STT3A complex (Ruiz-Canada et al., 2009). As very few posttranslational glycosylation sites have been described (Bolt et al., 2005; Bas et al., 2011; Tamura et al., 2011), we do not known why certain sequons are skipped by STT3A and then modified by STT3B, nor is it known whether glycosylation site skipping is a common or rare event.

One class of glycosylation site that might be prone to skipping by the STT3A isoform of the OST complex are NXT/S sites located near the extreme C terminus of a protein. In vitro experiments using ribosome-tethered nascent polypeptides indicate that a NXT/S site in a nascent polypeptide first becomes accessible to the OST active site when it is located 65–75 residues from the peptidyltransferase center on the ribosome (Whitley et al., 1996; Nilsson et al., 2003; Deprez et al., 2005). Movement of a nascent polypeptide past the OST active site is limited by the protein synthesis elongation rate, which is roughly six residues/second in mammalian cells (Hershey, 1991). After chain termination, sequons located in the last 65–75 residues of a protein might move by STT3A more rapidly and be skipped. Although these C-terminal glycosylation sites may be prone to skipping, modification of such sites can be critical for proper folding and subsequent vesicular transport of secretory proteins (Matzuk and Boime, 1988; McKinnon et al., 2010).

Glycosylation sites that are inserted into the C-terminal epitope tags of recombinant proteins are posttranslationally glycosylated when expressed in *Xenopus laevis* oocytes (Pult et al., 2011). However, it is not known which OST isoform is responsible for C-terminal glycosylation, nor is it known what factors influence the modification efficiency of naturally occurring C-terminal sites. A comprehensive biochemical and bioinformatics analysis of naturally occurring C-terminal glycosylation sites has not been previously reported. Our bioinformatics analysis of large glycopeptide databases from fungal, plant, and metazoan organisms indicates that the efficiency and mechanism of glycosylation of extreme C-terminal sites is evolutionarily conserved. Here, we show that extreme C-terminal glycosylation sequons in human proteins are glycosylated by an STT3B-dependent posttranslocational pathway. The kinetics and modification efficiency of C-terminal sites is strongly dependent on the acceptor site sequence. NXT sites are modified more rapidly and efficiently than NXS sites by STT3B even when both sites are present in the same polypeptide. Glycosylation sites that are located >50 residues from the C terminus show a reduced dependence on STT3B, indicating that 50 residues can be considered as a boundary for extreme C-terminal glycosylation sites.

Results

Bioinformatics analysis of C-terminal glycosylation sites

The most comprehensive list of experimentally verified N-glycosylation sites for a mammal has been obtained by glycoproteomic analysis of four tissues and plasma from mice (Zielinska et al., 2010). Based upon detection of previously documented murine glycopeptides, coverage of the murine glycoproteome may be as high as 75%. Using this large database of 4,922 high confidence murine glycosylation sites that match the N-(X≠P)-T/S/C sequon, we determined the percentage of modified sites that fall within each decile of protein length (Fig. 1 A). As observed previously with smaller glycopeptide datasets (Gavel and von Heijne, 1990; Ben-Dor et al., 2004), the N-terminal and C-terminal regions of murine glycoproteins have a lower density of N-linked glycans than internal regions.

Biological explanations for N-terminal or C-terminal reductions in glycan density are obscured by decile analysis because of the wide range in glycoprotein sequence length. To obtain a length-based assessment of glycan distribution, we calculated the normalized glycan density (glycans/1,000 amino acid residues) in 25-residue increments starting from the C terminus of the murine glycoproteins. The C-terminal regions of glycoproteins have 2.5-fold fewer N-linked glycans than internal regions (Fig. 1 B). The transition between the regions of lower and higher glycan density was surprisingly abrupt and occurred roughly 75 residues from the C terminus. The NXT/NXS ratio for the C-terminal glycopeptides was higher than for subsequent internal 75 residue blocks, each of which were similar to the bulk NXT/NXS ratio (Fig. 1 B, inset).

The density of total sequons (NXT/S/C) was reduced in the C-terminal 75 residues of the murine glycoproteins (Fig. 1 C). Color-coded dashed lines indicate the expected frequency for NXT, NXS, or NXC sequons based upon the amino acid composition of the murine glycoproteins. There is a strong positive selection for NXT sequons and a moderate selection for NXS sequons in murine glycoproteins. Examination of the C-terminal 75 residues revealed reduced positive selection for NXT sites as well as negative selection for NXS sites. NXC sites, which are modified at a 10-fold lower frequency than NXT sites, are subject to negative selection regardless of location. The apparent frequency of sequon modification (Fig. 1 D) was calculated by dividing the glycan density by the sequon density. Extreme C-terminal sequons are modified less frequently than internal sequons primarily because of a twofold reduction in the modification frequency of NXS sites (Fig. 1 D, inset).

STT3B-dependent glycosylation of extreme C-terminal sequons

Secretory proteins and soluble lysosomal proteins were the source of 52% of the glycopeptides located in the final 75 residues of murine glycoproteins. To address the mechanism of glycosylation of C-terminal sequons, we selected several secretory or lysosomal glycoproteins for analysis in HeLa cells. Sex hormone–binding globulin (SHBG), a 402 residue secreted glycoprotein, is an ideal substrate to investigate glycosylation of C-terminal sites (Fig. 2 A) because the two secreted forms of SHBG differ by having either one or two N-linked oligosaccharides (Bocchinfuso et al., 1992). HeLa cells transfected with expression vectors encoding wild-type SHBG or various SHBG glycosylation site mutants were pulse labeled and chased for 0 or 20 min. Wild-type SHBG could be resolved into two major polypeptides after pulse–chase labeling (Fig. 2 B). Removal of N-linked glycans by endoglycosidase H digestion yielded a
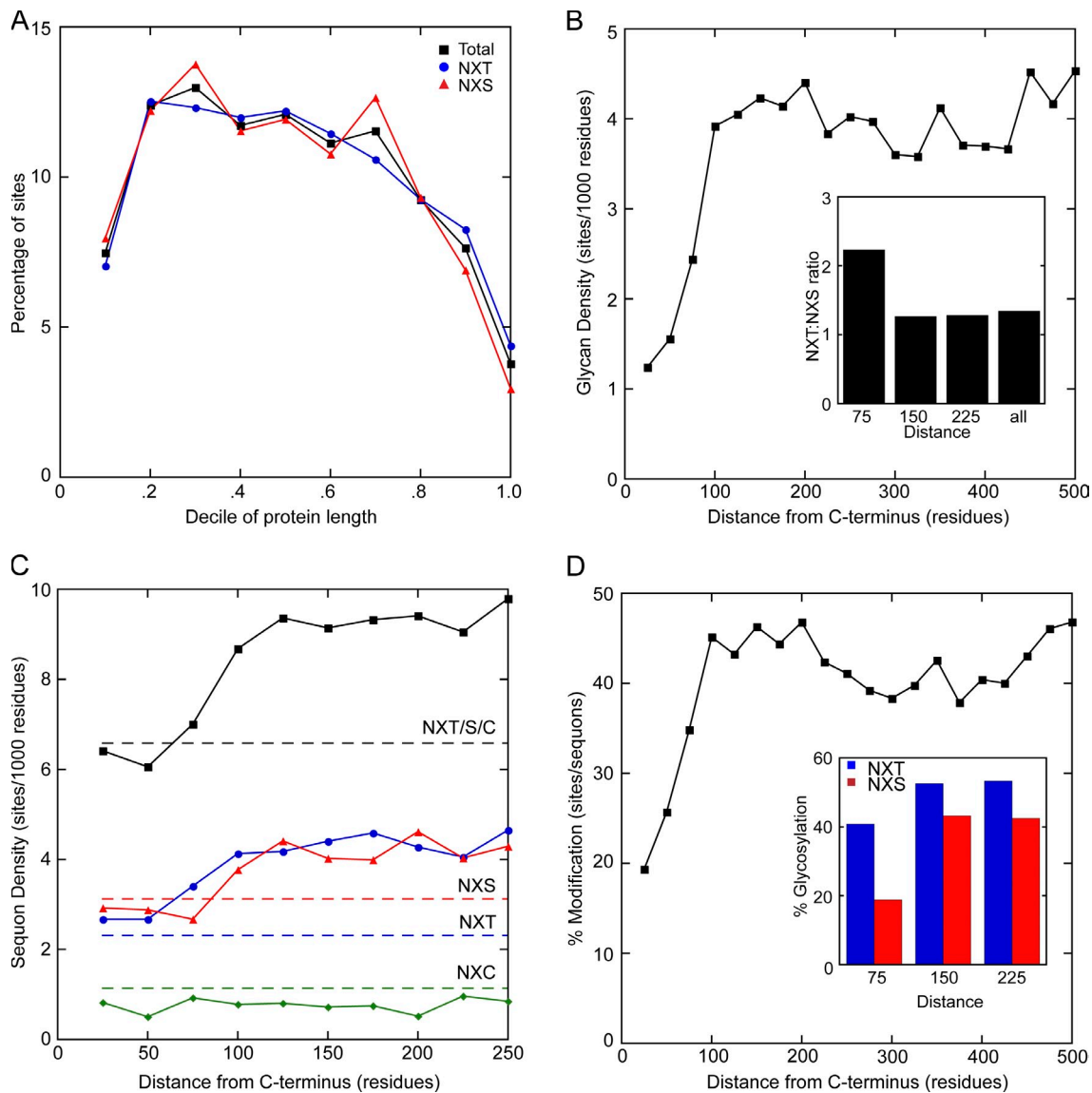
Figure 1. **Bioinformatics analysis of C-terminal glycosylation sites in murine glycoproteins.** The murine glycopeptide database consisted of 4,922 glyco-peptides derived from 1,902 proteins. (A) The location of glycans in murine glycoproteins distributed by decile of protein length. (B) The normalized glycan density versus distance in 25-residue increments from the C terminus. The NXT/NXS ratio for glycopeptides versus distance from the C terminus is shown in the inset. (C) The sequon density versus distance from the C terminus. Total sequons (NXT/S/C; black squares), NXT sites (blue circles), NXS sites (red triangles), and NXC sites (green diamonds) are shown. The 1,902 proteins had 13,326 sequons. Color-coded dashed lines are expected sequon densities for murine glycoproteins based upon the amino acid composition of the proteins. (D) Apparent modification efficiency versus distance from the C terminus. The inset shows the apparent modification efficiency for NXT and NXS sites in 75 residue increments from the C terminus.

single more rapidly migrating polypeptide, confirming that the SHBG doublet is caused by heterogeneous glycosylation. Most of the SHBG chains have a single glycan at the conclusion of the 3-min pulse-labeling period, whereas 50% of SHBG chains acquire a second glycan during the 20-min chase period. Trim-ming of glucose residues from N-linked glycans during the 20-min chase incubation is responsible for the minor increase in gel mobility of monoglycosylated SHBG that occurs during the chase incubation (Fig. 2 B, compare 0- and 20-min chase sam-ples). The two single-site SHBG mutants (N380Q and N396Q) were pulse labeled to determine which site was glycosylated during each phase of the labeling experiment. The $N_{396}GT$ site was efficiently modified, primarily during the initial pulse (Fig. 2 B, N380Q mutant). The $N_{380}RS$ site was incompletely

modified (0.7 glycan/site), primarily during the chase period (Fig. 2 B, N396Q mutant). Additional mutants were analyzed to determine why the kinetics and modification efficiency of the $N_{380}RS$ and $N_{396}GT$ site differed. Converting the $N_{380}RS$ into an $N_{380}RT$ site (S382T and S382T N396Q mutants) increased both the glycosylation rate and extent of site occupancy. Conversely, replacing threonine with serine in the $N_{396}GT$ site (T398S and T398S N380Q mutants) decreased the rate and extent of gly-cosylation. These differences in sequon modification (NXT vs. NXS) are consistent with the bioinformatics analysis, indicating that C-terminal NXS sites are modified less efficiently than in-ternal NXS sites or C-terminal NXT sites (Fig. 1 D, inset).

Wild-type SHBG and the two single-site mutants were ex-pressed in HeLa cells that had been transfected with previously
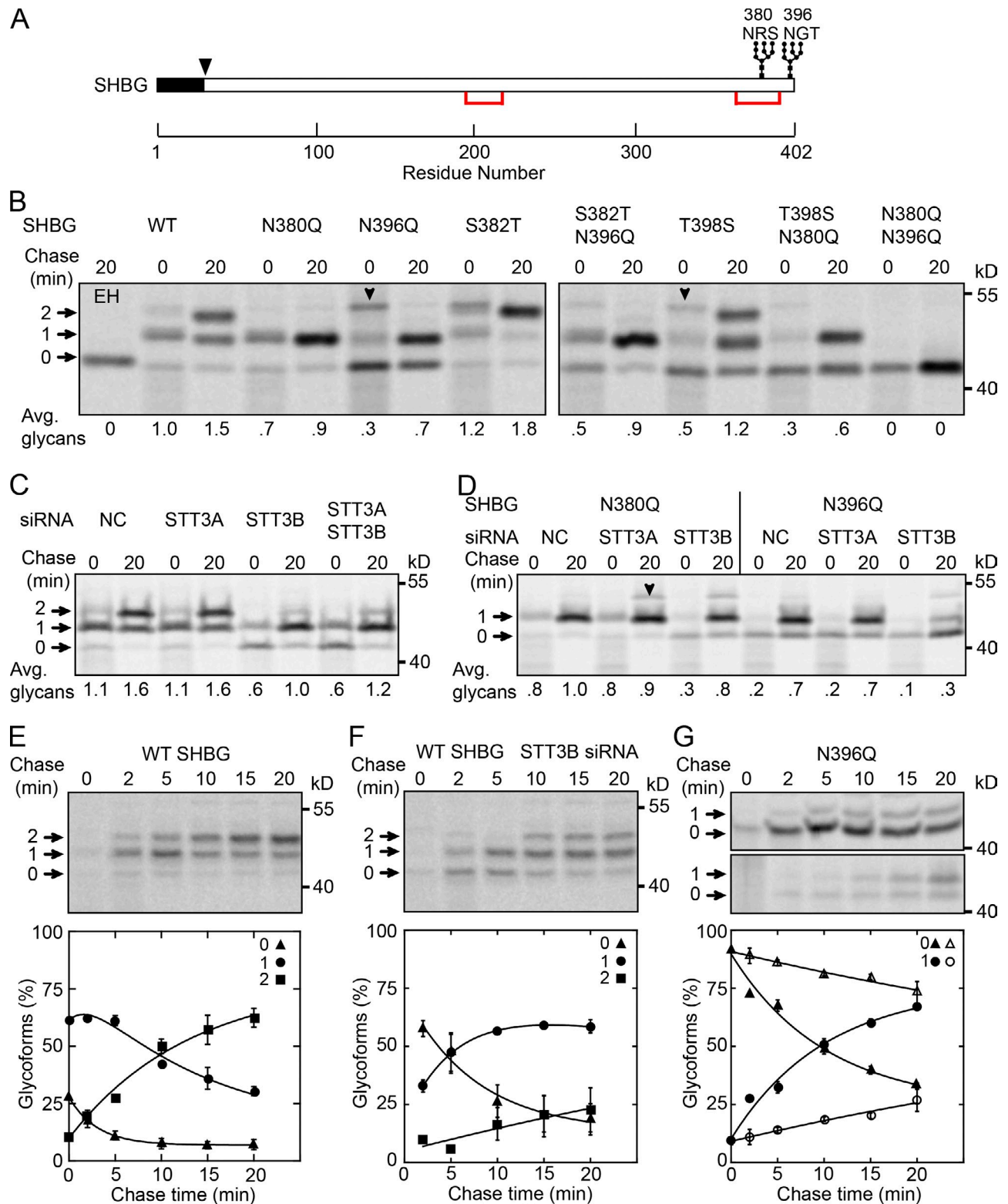
Figure 2. **Posttranslocational glycosylation of SHBG.** (A) Diagram of SHBG showing the signal sequence (black), cleavage site (arrowhead), glycosylation sites, and disulfide bonds (red lines). (B–G) HeLa cells were treated with no siRNA (B), negative control (NC) siRNA, or siRNAs specific for STT3A or STT3B as indicated (C and D) for 48 h before transfection with SHBG expression vectors. Cells were pulse labeled for 3 min (B), 5 min (C), or 2 min (D–G) and chased as indicated. EH designates digestion with endoglycosidase H. SHBG glycoforms (0–2 glycans) were immunoprecipitated with anti-SHBG sera and resolved by SDS-PAGE. Downward pointing arrowheads (B and D) designate a nonspecific immunoprecipitation product. Values listed below gel lanes (B–D) are the mean of two determinations. (E–G) Pulse–chase analysis of wild-type SHBG (E and F) or the N396Q SHBG mutant (G) in cells treated with NC siRNA (E and G, top image) or in cells treated with the STT3B siRNA (F and G, bottom image). The percentage of SHBG molecules that have zero to two glycans was calculated. Open symbols in G designate STT3B-depleted cells. Symbols in E–G are the mean of two determinations; error bars indicate individual data points. Mean values were fit to single or double exponential equations as described in Materials and methods. The indicated molecular masses shown in this and subsequent figures are of prestained molecular mass markers electrophoresed on all gels. Avg., average; WT, wild type.

characterized siRNAs that are specific for the *STT3A* or *STT3B* mRNAs (Ruiz-Canada et al., 2009). A10-fold depletion of STT3A did not reduce glycosylation of wild-type SHBG or either single-site mutant relative to a negative control siRNA (Fig. 2, C and D). siRNA-mediated depletion of STT3B, which lowers STT3B levels by roughly sixfold (unpublished data; Ruiz-Canada et al., 2009), reduced glycosylation of wild-type SHBG and both single-site mutants. Interestingly, the suboptimal $N_{380}RS$ site was more sensitive than the $N_{396}GT$ site to the reduction in STT3B levels (Fig. 2 D), likely because of the different kinetics of modification. Simultaneous depletion of both STT3A and STT3B was similar to STT3B depletion alone, indicating that STT3A has little, if any, role in glycosylation of SHBG (Fig. 2 C).

A more extensive pulse–chase analysis of SHBG glyco-sylation was conducted to determine the $t_{1/2}$ for glycosylation of the C-terminal sites (Fig. 2 E). Quantification of the pulse–chase data yielded the expected relationship between a nonglyco-sylated precursor, a monoglycosylated intermediate, and the fully glycosylated product. Disappearance of nonglycosylated SHBG, which corresponds to glycosylation of the $N_{396}GT$ site occurred with a $t_{1/2}$ of 2 ± 1 min. Accumulation of diglycosyl-ated SHBG, which is diagnostic of glycosylation of the $N_{380}RS$ site, occurred with a $t_{1/2}$ of 9 ± 2 min. The N396Q mutant was analyzed to determine whether the kinetics of glycosylation of the $N_{380}RS$ site was influenced by glycosylation of the $N_{396}GT$ site (Fig. 2 G, top and bottom, black symbols). The $t_{1/2}$ for gly-cosylation of the $N_{380}RS$ site in the N396Q mutant was 7 ± 2 min, indicating that glycosylation of the two sites is indepen-dent. Depletion of STT3B reduced the rate and final extent of glycosylation of both acceptor sites in wild-type SHBG (Fig. 2 F). The calculated $t_{1/2}$ values for glycosylation of the $N_{396}GT$ site and the $N_{380}RS$ site in STT3B-depleted cells increased to 5 and >30 min, respectively. Depletion of STT3B reduced the rate of glycosylation of the $N_{380}RS$ site by more than fourfold in the N396Q mutant (Fig. 2 G, open symbols). Collectively, these experiments indicate that the extreme C-terminal sites in SHBG are glycosylated by STT3B after the complete polypeptide has passed through the protein translocation channel; hence, modi-fication of these sites is obligatorily posttranslational and mech-anistically posttranslocational.

Internal glycosylation sites enhance C-terminal glycosylation of β-glucuronidase (β-GUS)

Our analysis of the murine glycopeptide database indicated that most glycoproteins with an extreme C-terminal glycan have one or more internal NXT/S sequons. Human β-GUS has three internal sites (sequons 1–3) and one C-terminal site (Fig. 3 A, sequon 4). Using a Myc-DDK–tagged β-GUS construct, we confirmed a previous observation (Shipley et al., 1993) that mutation of two or three internal NXT/S sites (β-GUSΔ23 or β-GUSΔ123) resulted in the accumulation of glycoform dou-blets, indicating incomplete modification of N631 (Fig. 3 B). These doublets were prominent in the 0-min chase samples (β-GUSΔ23 and β-GUSΔ123) unless the $N_{631}ET$ site (β-GUSΔ4) was eliminated (Fig. 3 B). β-GUS chains lacking 1–4 glycans were synthesized by HeLa cells when STT3A and STT3B were

simultaneously depleted (Fig. 3 C). Depletion of STT3A alone did not cause hypoglycosylation of wild-type β-GUS or either mutant, whereas depletion of STT3B resulted in glycoform doublet accumulation for wild-type β-GUS and the β-GUSΔ123 mutant (Fig. 3 C). These results indicate that the three internal sequons in β-GUS can be glycosylated by either OST isoform, whereas STT3B is necessary for efficient glycosylation of the C-terminal $N_{631}ET$ site.

In vivo biosynthesis of β-GUS takes roughly 2 min. A ki-netic analysis of wild-type β-GUS glycosylation showed that 70% of the translocated β-GUS chains were fully glycosylated by the end of a 5-min pulse-labeling period (Fig. 3 D). The re-maining 30% of β-GUS chains, which lack a single glycan, disappeared during the chase. The nonglycosylated band that co-migrates with the endoglycosidase H digestion product is a nontranslocated precursor that was excluded from the kinetic analysis. Quantification of the pulse–chase experiment (Fig. 3 E, squares) yielded a $t_{1/2}$ of 1.9 ± 0.2 min for posttranslocational glycosylation of wild-type β-GUS. Pulse–chase analysis of the β-GUSΔ123 mutant showed that the $N_{631}ET$ site was glycosyl-ated at a reduced rate ($t_{1/2}$ = 7.6 ± 4 min) when the internal sites were absent (Fig. 3 E, circles). Although STT3B depletion strongly inhibited glycosylation of the $N_{631}ET$ site (Fig. 3 D), a $t_{1/2}$ could not be calculated without making an assumption con-cerning the final extent of modification (Fig. 3 E, triangles). Glycosylation of β-GUS mutants with one internal site (Δ12, Δ13, or Δ23) is sensitive to STT3B depletion (Fig. 3 F). Pulse–chase experiments analyzed as in Fig. 3 E showed that a single internal glycan, regardless of position (Δ12, Δ13, or Δ23), re-duced the $t_{1/2}$ for glycosylation of the C-terminal $N_{631}ET$ site relative to the β-GUS Δ123 mutant (Fig. 3 G).

C-terminal glycosylation of type II membrane proteins

Type II ($N_{cytosol}$-$C_{lumenal}$) membrane proteins are the second most abundant class of murine proteins with glycans in their C-terminal 75 residues (22% of C-terminal glycopeptides). Human CD40 ligand was selected as an example of a type II membrane pro-tein with a single C-terminal glycosylation site (Fig. 4 A). Gly-cosylation of the $N_{240}VT$ site was reduced slightly by STT3B depletion in the Myc-DDK– and MycΔDDK-tagged versions of CD40L (Fig. 4 B). Replacing the NVT sequon with an NVS se-quon (T242S mutant) reduced the modification frequency of the site and caused an increase in sensitivity to STT3B depletion.

CD69 has a C-terminal NVT site and an internal NAC site (Fig. 4 C). Although depletion of either STT3A or STT3B caused reductions in glycosylation of CD69-Myc-DDK, it was not clear which site or sites were affected (Fig. 4 D). Elimina-tion of the NAC site (N111Q mutant) confirmed a previous study that heterogeneous glycosylation of CD69 can be ascribed to incomplete modification of the $N_{111}AC$ site (Vance et al., 1997). In the context of the shorter epitope tag (MycΔDDK), glycosylation of the $N_{166}VT$ site was mildly sensitive to STT3B depletion but not to STT3A depletion. Replacing the NVT site with an NVS site (T168S mutant) reduced the glycosylation of N166. Thus, extreme C-terminal sequons in type II membrane proteins are also STT3B substrates.
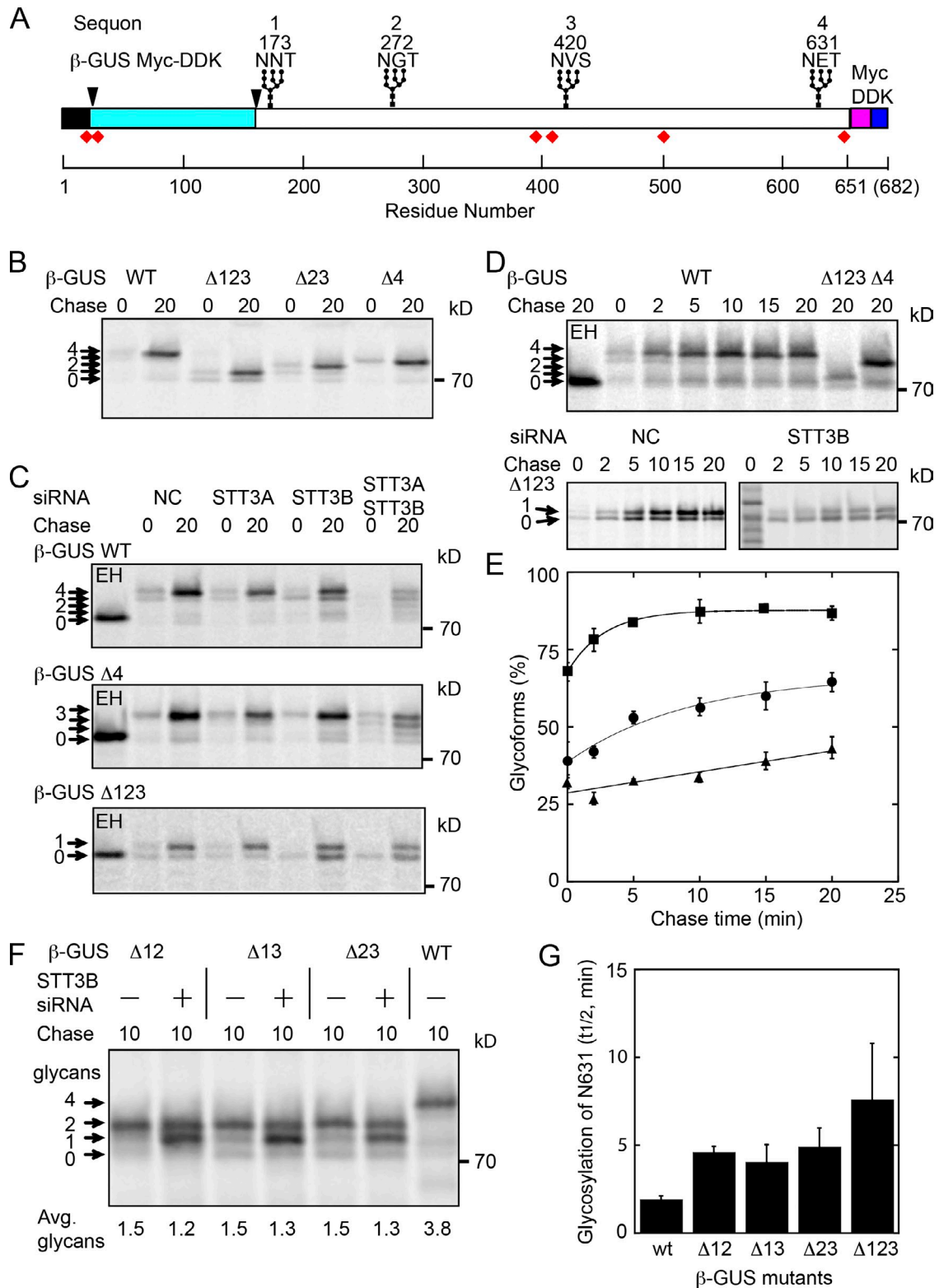
Figure 3. **Glycosylation of β-GUS in siRNA-treated cells.** (A) Diagram of β-GUS showing the signal sequence (black), proteolytic cleavage sites (arrowheads), glycosylation sites, free cysteine residues (red diamonds), and a Myc-DDK tag. The four sequons are numbered 1–4; β-GUS mutants lacking one or more sequons are designated as β-GUSΔXYZ, in which XYZ is the list of mutated sequons. (B–D) Transfected HeLa cells were pulse labeled for 5 min and chased as indicated with (C, D, and F) or without (B) prior treatment with siRNAs. Anti-DDK immunoprecipitates were analyzed by SDS-PAGE to resolve β-GUS with zero to four N-linked glycans. (D) Pulse–chase analysis of glycosylation of the $N_{631}ET$ site in wild-type (WT) β-GUS (top) or β-GUSΔ123 (bottom). (E) The percentages of wild-type β-GUS chains that have four glycans (squares) or β-GUSΔ123 chains that have one glycan (circles and triangles) are shown. Triangles designate cells treated with siRNA for STT3B. Symbols in E are the mean of two determinations; error bars indicate individual data points. (F) Values shown below F are the mean of two determinations. (G) The kinetics ($t_{1/2}$ values) of glycosylation of the $N_{631}ET$ site in the Δ12, Δ13, and Δ23 mutants was determined by pulse–chase analysis as in D and E. The values shown are the $t_{1/2}$ values obtained by fitting the experimental data to a single exponential equation; error bars designate standard deviations. Avg., average; EH, endoglycosidase H; NC, negative control.
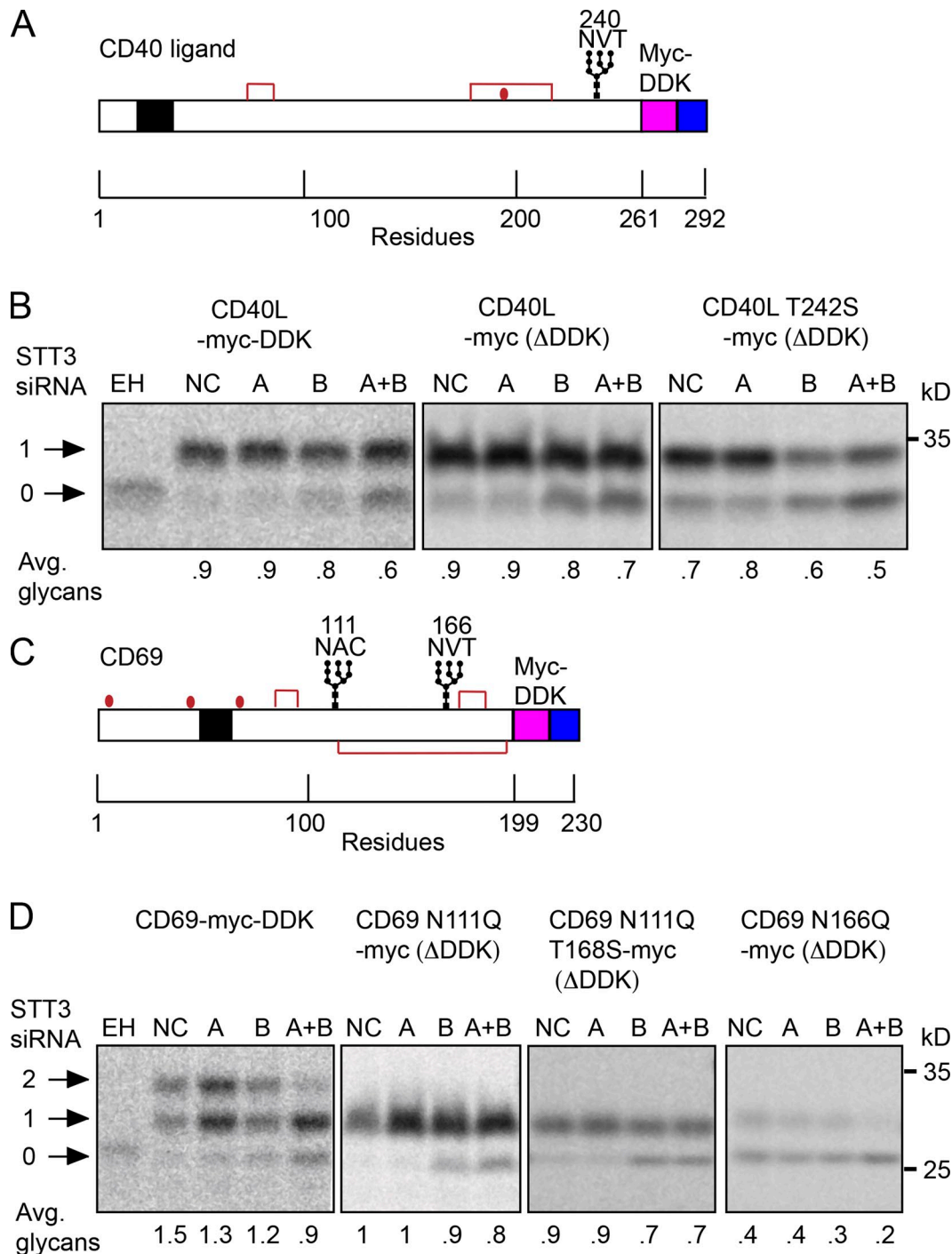
Figure 4. **Glycosylation of type II transmembrane proteins in siRNA-treated cells.** (A and C) Diagram of the CD40 ligand (A) and CD69 (C) showing the transmembrane domain (black), glycosylation sites, disulfide bonds (red lines), and free cysteines (red dots). (B and D) HeLa cells cotransfected with the indicated plasmids and siRNAs were pulse labeled for 2 min (B) or 4 min (D) and chased for 20 min. Polypeptides with zero to two N-linked glycans were resolved by SDS-PAGE. Values listed below B and D are the mean of two determinations. Avg., average; EH, endoglycosidase H; NC, negative control.

Another role for the STT3B isoform of the OST is to glycosylate internal sites that are skipped by STT3A. We constructed a CD69 N166Q-MycΔDDK mutant to determine which OST isoform mediates glycosylation of the $N_{111}AC$ site (Fig. 4 D). Depletion of either STT3A or STT3B caused a partial reduction in $N_{111}AC$ modification, indicating that both STT3A and STT3B contribute to the observed, albeit partial, glycosylation of the suboptimal $N_{111}AC$ site.

Defining a boundary for C-terminal glycosylation

The serum glycoprotein transferrin (Tf) has one internal and one C-terminal glycosylation site (Fig. 5 A). Hypoglycosylated Tf is a diagnostic marker for the family of diseases known as congenital disorders of glycosylation type 1 (CDG-1). Most subtypes of CDG-1 are caused by incomplete assembly of dolichol pyrophosphate–linked $GlcNAc_2Man_9Glc_3$, the oligosaccharide

donor for N-glycosylation (Haeuptle and Hennet, 2009). Mass spectroscopy of serum Tf from CDG-1 patients has indicated that the C-terminal site is more frequently skipped than the internal site (Hülsmeier et al., 2007).

HeLa cells transfected with DDK-His–tagged Tf were subjected to pulse–chase analysis to determine whether the C-terminal site in Tf undergoes posttranslocational glycosylation (Fig. 5 B). Immunoprecipitation of the pulse-labeled samples with anti-DDK sera indicated that Tf was fully glycosylated at the end of the 5-min pulse. No evidence for posttranslocational modification of either sequon was obtained by pulse–chase analysis of the two single-site Tf mutants.

Depletion of STT3A plus STT3B resulted in synthesis of Tf-DDK-His chains that lack one to two N-linked glycans (Fig. 5 C). Monoglycosylated Tf was also produced when STT3A was depleted but absent when STT3B was depleted. Analysis of the N630Q mutant indicated that the $N_{432}KS$ site was particularly sensitive to STT3A depletion. A subset of the siRNA depletion experiments were conducted using an untagged Tf construct (Fig. 5 D) to determine whether STT3B is required for glycosylation of the $N_{630}VT$ site when this site is located closer (68 residues) to the C terminus. As observed for the DDK-His–tagged construct, wild-type Tf was sensitive to STT3A depletion. The C-terminal $N_{630}VT$ site (N432Q mutant) was relatively insensitive to depletion of either STT3A or STT3B.

To define a boundary for STT3B-dependent posttranslocational glycosylation, additional NVT sequons were inserted into the Tf-DDK-His construct or the N432Q Tf-DDK-His construct. The I-1, I-2, and I-3 sequons are located 84, 64, and 14 residues from the C terminus of Tf-DDK-His, respectively (Fig. 5 A). The I-1 and I-2 sites were more efficiently glycosylated than the I-3 site (Fig. 5 E). Depletion of STT3B eliminated glycosylation of the I-3 site but did not reduce glycosylation of the I-1 or I-2 sites (Fig. 5 F). Truncation of the DDK-His tag by 11 residues (Tf I-2ΔHis) reduced the distance between the inserted I-2 site and the C terminus to 53 residues. As the glycosylated form of Tf I-2ΔHis co-migrated with wild-type Tf and was insensitive to STT3B depletion (Fig. 5 G), it was initially unclear whether the inserted site had been modified. Limited digestion of the Tf I-2 and Tf I-2ΔHis products with endoglycosidase H produced a glycan ladder that allows facile counting of N-linked glycans (Fig. 5 H). Both the Tf I-2 and Tf I-2ΔHis product have three N-linked glycans consistent with efficient modification of the inserted NVT site.

To obtain a better understanding of the boundary for posttranslocational C-terminal glycosylation, the location of murine glycopeptides was plotted as a running sum versus distance from the C terminus (Fig. 5 I, top plot). Remarkably, the plotted data could be analyzed as two intersecting lines, the slopes of which are proportional to the observed glycan density. The calculated intersection point for the two lines is 66 residues, which falls within the range of reported distances in amino acid residues between the peptidyltransferase site on the ribosome and the OST active site (Whitley et al., 1996; Deprez et al., 2005). The location and STT3B dependence of the C-terminal glycosylation sites in SHBG, β-GUS Myc-DDK, Tf, Tf-DDK-His, CD40L, CD69, blood coagulation factor VII, and prosaposin

are plotted on the same abscissa (Fig. 5 I, bottom plot). With one noteworthy exception (Fig. 5 I, factor VII, inverted black triangle), sequons that show dependence upon STT3B fall on the low glycan density arm of the glycopeptide versus distance plot. STT3B-independent sites, including those in Tf and prosaposin, are located on the high glycan density arm of the plot. The I-2ΔHis and I-2 sites in Tf fall in an intermediate region near the intersection point where the observed glycan density begins to exceed the linear regression fit for the low-density arm of the plot (Fig. 5 I, gray square and gray circle).

C-terminal glycosylation in model eukaryotes

Recently, N-linked glycopeptides from six diverse model organisms were identified by mass spectrometry (Zielinska et al., 2012). These databases were analyzed as in Fig. 5 I to determine the density of sequons and glycans in the C-terminal 150 residues of glycoproteins (Fig. 6, glycans are blue lines, and sequons are red lines). Sequon and glycan densities that were calculated by linear regression analysis are shown adjacent to each line or pair of intersecting lines. With the exception of *Schizosaccharomyces pombe*, there is a reduced density of N-glycans in the C-terminal 75 residues of the glycoproteins. Reductions in C-terminal glycan density ranged between 1.4-fold for *Saccharomyces cerevisiae* to 2.3–2.5-fold for *Drosophila melanogaster*, *Danio rerio*, and *Mus musculus*. With the exception of *Caenorhabditis elegans*, the reduced C-terminal glycan density is in part explained by a reduced density of C-terminal sequons (Fig. 6). We next asked whether NXS sequons, NXT sequons, or both were reduced in C-terminal segments of glycoproteins (Table 1 and Table 2). For all organisms except *S. pombe* and *C. elegans*, positive selection of NXT sequons is reduced near the C terminus (Table 1), whereas NXS sequons are subject to negative selection (Table 2). As observed for *M. musculus* (Fig. 1 D), the apparent modification frequency of NXT sequons shows little or no reduction in the C-terminal 75 residues (Table 1). The modification frequency for C-terminal NXS sequons was reduced in all organisms except *S. pombe* and *Arabidopsis thaliana* (Table 2). Thus, the overall reduction in C-terminal glycan density in glycoproteins from metazoan organisms is caused by a reduction in sequon density and a lower modification efficiency of C-terminal NXS sites (Table 2).

The number of eukaryotic STT3 sequences in the protein sequence database has increased roughly fivefold since we last examined the relationship between eukaryotic STT3 proteins (Kelleher and Gilmore, 2006). Using this larger dataset, it is now clear that the genomes of all sequenced metazoan organisms, with the exception of species in the genus *Caenorhabditis*, encode both STT3A and STT3B (Fig. S1). The genomes of all multicellular plants and some unicellular plants also encode two STT3 proteins. One clade of plant STT3 proteins is more closely related to metazoan STT3A proteins, whereas the second clade is related to protist STT3 proteins. Fungal genomes encode a single STT3 protein. Metazoan STT3B proteins, which include the STT3 proteins of *Caenorhabditis* species, are more closely related to fungal STT3 proteins. Thus, the three model organisms (*S. cerevisiae*, *S. pombe*, and *C. elegans*) that show
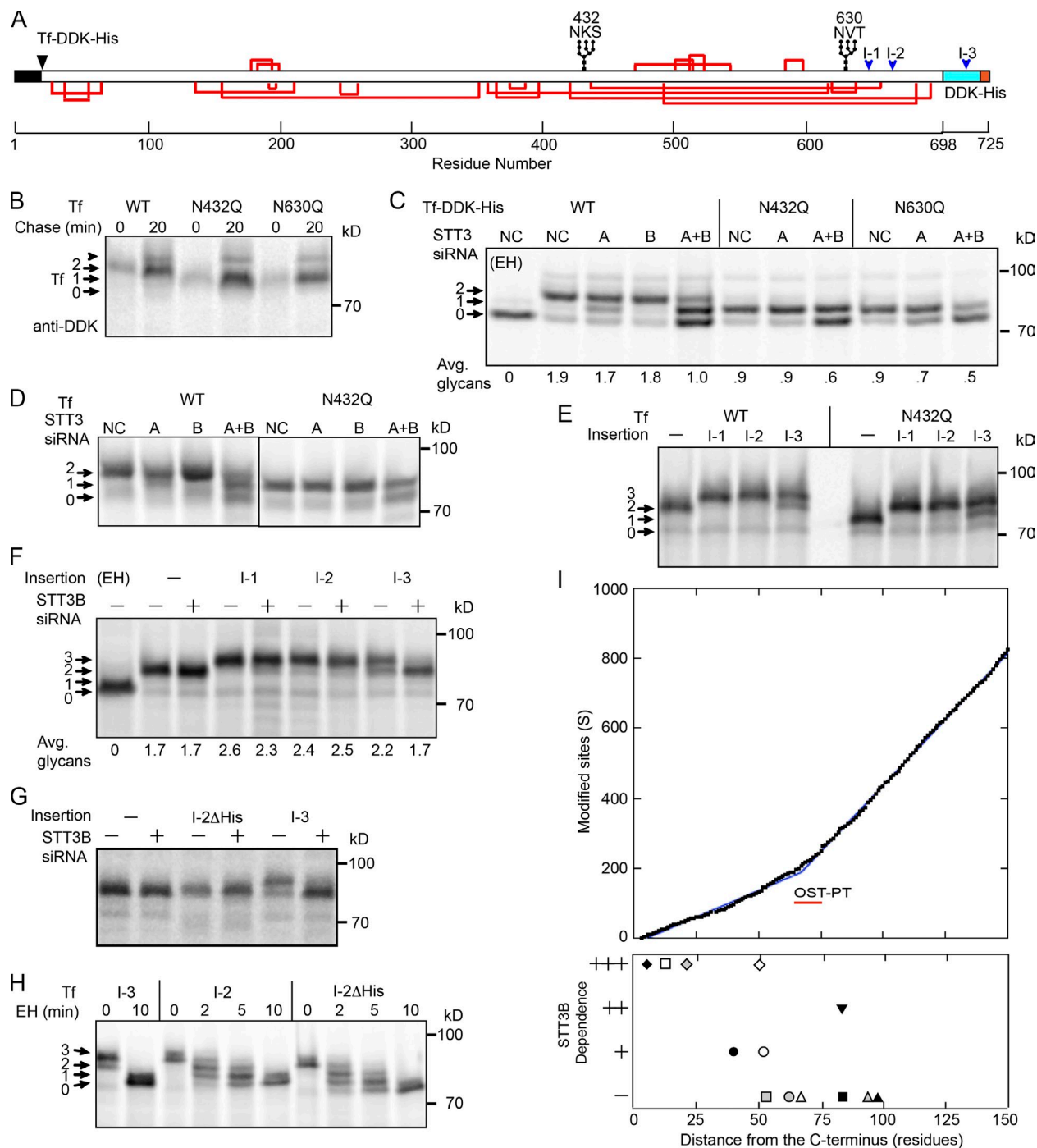
Figure 5. **Glycosylation of Tf in siRNA-treated cells.** (A) Diagram of transferrin (Tf) showing the signal sequence (black), glycosylation sites, disulfide bonds (red lines), NVT sequon insertion sites (I-1, I-2, and I-3), and C-terminal DDK-His tag. (B–H) Tf-DDK-His was immunoprecipitated with the anti-DDK antibody (B, C, and E–H), whereas untagged Tf was immunoprecipitated with the anti-Tf antibody (D). (B) HeLa cells transfected with DDK-His–tagged wild-type Tf or Tf single-site mutants (N432Q or N630Q) were pulse labeled for 5 min and chased as indicated. (C and D) Pulse–chase labeling (5-min pulse and 20-min chase) of DDK-His–tagged Tf (C) or untagged Tf (D) in siRNA-treated HeLa cells. Black line indicates that intervening lanes have been spliced out. (E–H) DDK-His–tagged Tf constructs that have NVT insertions were pulse labeled for 10 min with (F and G) or without (E and H) prior treatment of the cells with STT3B siRNA. (H) Immunoprecipitated samples were subjected to digestion with endoglycosidase H (EH) for brief periods as indicated to obtain digestion intermediates that were resolved by SDS-PAGE. Values listed below gel lanes (C and F) are the means of two determinations. (I) Murine glycopeptides derived from 1,902 proteins are plotted as a running sum (black squares) relative to distance from the C terminus. Linear regression fits (blue lines) of the two arms of the curve intersect at a length of 66 residues. Estimates of STT3B dependence of glycosylation of SHBG $N_{396}GT$ (black diamond), SHBG $N_{380}RS$ (gray diamond), Tf I-3 (open square), β-GUS-Myc-DDK $N_{631}ET$ (open diamond), Tf I-2ΔHis (gray square), Tf I-2 (gray circle), untagged Tf $N_{630}VT$ (open triangle), Tf I-1 (black square), Tf-DDK-His $N_{630}VT$ (gray triangle), CD40L-MycΔDDK (open circle), CD69-MycΔDDK (black circle), factor VII $N_{360}IT$ (inverted black triangle), and prosaposin $N_{429}ST$ (black triangle) are shown. Our scaling system is qualitative because STT3B depletions are incomplete. The STT3B dependence values are defined as follows based upon the observed reductions in sequon modification in STT3B-depleted cells: >50% reduction (+++, strong); 25–50% reduction (++, moderate); 5–25% reduction (+, detectable); and 0–5% reduction (−, independent). Values for factor VII $N_{360}IT$ and prosaposin $N_{429}ST$ are taken from Ruiz-Canada et al. (2009). The red bar (OST-PT) designates the reported distance in residues between the OST active site and the ribosomal peptidyltransferase center. Avg., average; NC, negative control; WT, wild type.
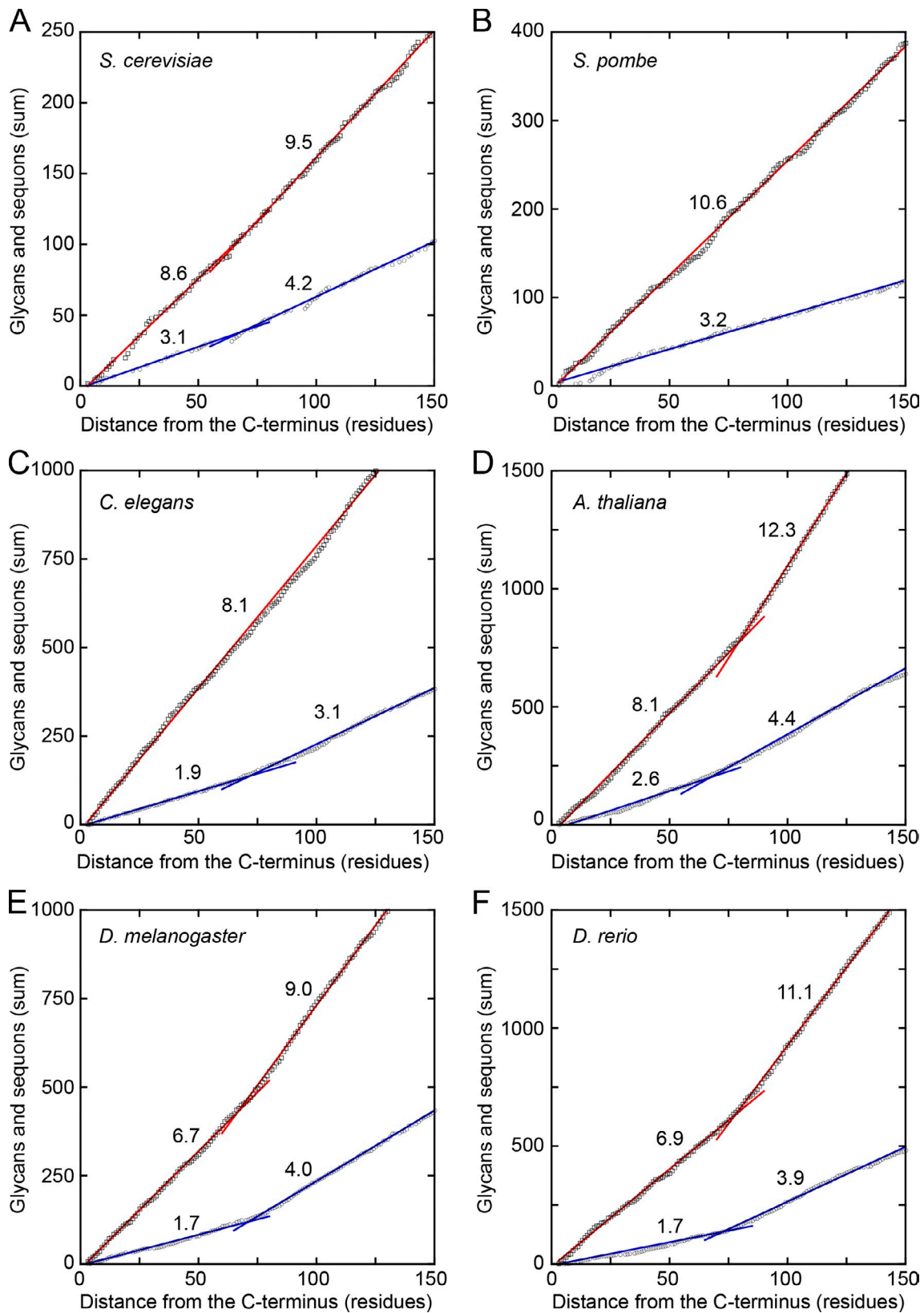
Figure 6. **Glycan and sequon densities in C-terminal regions of glycoproteins from model organisms.** (A–F) The location of glycopeptides (open circles, blue lines) and sequons (open squares, red lines) in glycoproteins are plotted as running sums versus distance from the C terminus. The number of analyzed glycoproteins (P), high confidence glycopeptides (GP), and predicted sequons (S) for each organism is as follows: *S. cerevisiae* (168 P, 485 GP, and 1,337 S); *S. pombe* (245 P, 416 GP, and 1,543 S); *C. elegans* (988 P, 1,736 GP, and 5,310 S); *A. thaliana* (1,267 P, 2,164 GP, and 7,503 S); *D. melanogaster* (999 P, 2,208 GP, and 6,604 S); and *D. rerio* (1,193 P, 2,160 GP, and 7,810 S). In most cases, two lines, intersecting at 70 ± 5 residues, provided a superior fit compared with a single line. Numerical values adjacent to lines are sequon and glycan densities (sequons or glycans/1,000 residues) calculated from the slopes of the lines. Intersecting lines are extended to highlight the intersection point.

Table 1. **NXT acceptor sites in eukaryotic glycoproteins**

Organism	Sequons (observed/expected)		Glycan density	Apparent modification frequency	
	T-sequon	**C-sequon**	**C-glycan/T-glycan**	**T-glycan/T-sequon**	**C-glycan/C-sequon**
				%	%
S. cerevisiae	1.8	1.3	0.78	42	47
S. pombe	1.6	1.7	1.15	34	38
C. elegans	1.6	1.5	0.82	38	31
A. thaliana	2.0	1.4	0.73	32	34
D. melanogaster	1.6	1.3	0.74	39	38
D. rerio	1.7	1.2	0.77	31	33
M. musculus	1.9	1.2	0.56	44	37

An observed/expected ratio >1 indicates positive selection for NXT sequons. T-sequon, total NXT sequons; C-sequon, NXT sequons in the C-terminal region (70 ± 5 residues). The C-sequon density (NXT sequons/1,000 residues) for the C-terminal region was calculated from linear regression fits of plots of NXT sequons versus C-terminal distance. Expected sequon densities were calculated using the amino acid composition of the glycoproteins from each species. The density (glycans/1,000 residues) of NXT glycans in the C-terminal segment (70 ± 5 residues) of proteins was determined by linear regression of plots of NXT glycopeptides versus C-terminal distance. C-glycans, C-terminal glycans; T-glycans, total glycans. The apparent modification frequency was calculated by dividing the glycan density by the sequon density. The apparent modification frequency underestimates the actual percent modification caused by incomplete detection of all glycopeptides by mass spectrometry.

the least reduction in C-terminal glycan density express a single OST catalytic subunit that is more closely related to the STT3B isoform responsible for posttranslocational glycosylation of extreme C-terminal sites in human cells.

Discussion

Hypoglycosylation can cause defects in glycoprotein folding, secretion, and function that contribute to the pathophysiological symptoms displayed by patients that have congenital disorders of glycosylation (Haeuptle and Hennet, 2009). Although extreme C-terminal sequons are prone to be skipped by the translocon-associated STT3A isoform of the OST, these sequons are often evolutionarily conserved (e.g., N396 in SHBG and N631 in β-GUS), indicating that glycans at these positions are important. The α subunit of human chorionic gonadotropin (Matzuk and Boime, 1988) and von Willebrand factor (McKinnon et al., 2010) are examples of proteins that have extreme C-terminal glycans that are necessary for efficient protein folding and secretion. Here, we have presented evidence that extreme C-terminal sequons in proteins are glycosylated by a STT3B-dependent

posttranslocational pathway. Our analysis of large glycoprotein databases has revealed that C-terminal sequons are not uncommon and that C-terminal NXT sites appear to be glycosylated as efficiently as internal NXT sites.

Posttranslocational glycosylation of C-terminal acceptor sites

Because of the distance between the peptidyltransferase site and the OST active site, sequons in the last 65–75 residues of a protein are not glycosylated before chain termination (Fig. 7 b). Pulse–chase labeling of two secretory proteins (SHBG and β-GUS) indicated that C-terminal sites located in the last 50 residues of a protein are prone to be skipped by the translocation channel–associated STT3A isoform of the OST complex, hence their glycosylation was insensitive to STT3A depletion. Glycosylation of extreme C-terminal NXT sites was rapid ($t_{1/2}$ of ∼2 min) and efficient, whereas glycosylation of the NXS site in SHBG was roughly fourfold slower and did not reach completion. Three important conclusions can be drawn from the pulse–chase analysis of SHBG glycosylation. The order in which the sequons are modified in wild-type SHBG is not determined by

Table 2. **NXS acceptor sites in eukaryotic glycoproteins**

Organism	Sequons (observed/expected)		Glycan density	Apparent modification frequency	
	T-sequon	**C-sequon**	**C-glycan/T-glycan**	**T-glycan/T-sequon**	**C-glycan/C-sequon**
				%	%
S. cerevisiae	0.9	0.7	0.51	27	18
S. pombe	1.0	1.0	1.08	20	22
C. elegans	0.7	0.7	0.63	25	16
A. thaliana	1.1	0.9	0.78	25	26
D. melanogaster	1.0	0.8	0.37	26	13
D. rerio	1.0	0.9	0.56	24	15
M. musculus	1.3	0.9	0.34	36	17

An observed/expected ratio <1 indicates negative selection for NXS sequons. T-sequon, total NXS sequons; C-sequon, NXS sequons in the C-terminal region (70 ± 5 residues). The C-sequon density (NXS sequons/1,000 residues) for the C-terminal region was calculated from linear regression fits of plots of NXS sequons versus C-terminal distance. Expected sequon densities were calculated using the amino acid composition of the glycoproteins from each species. The density (glycans/1,000 residues) of NXS glycans in the C-terminal segment (70 ± 5 residues) of proteins was determined by linear regression of plots of NXS glycopeptides versus C-terminal distance. C-glycans, C-terminal glycans; T-glycans, total glycans. The apparent modification frequency was calculated by dividing the glycan density by the sequon density. The apparent modification frequency underestimates the actual percent modification caused by incomplete detection of all glycopeptides by mass spectrometry.
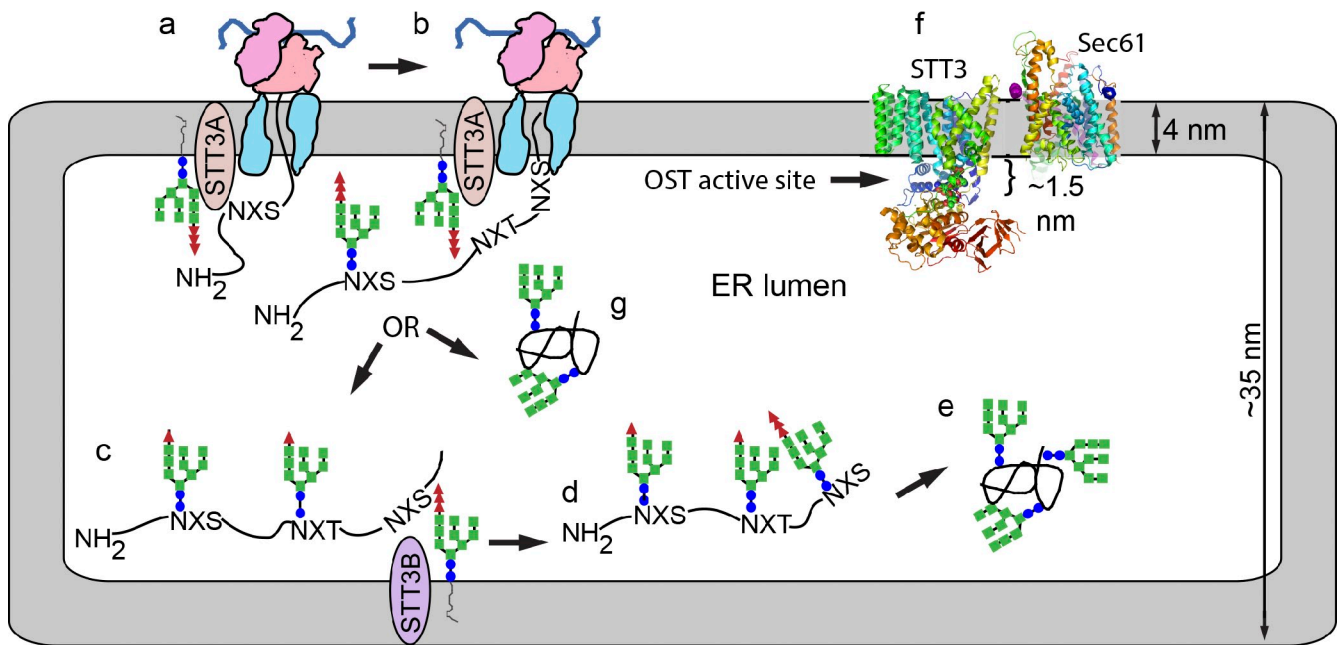
Figure 7. **Posttranslocational glycosylation of extreme C-terminal sequons.** (a) Co-translational glycosylation of internal sequons by the STT3A isoform of the OST. (b) Extreme C-terminal sequons (C-terminal 50–55 residues) are inaccessible to STT3A before chain termination and are frequently skipped by the STT3A complex. (c–e) Skipped sequons (c) can be recognized by the STT3B complex and glycosylated (d) before protein folding (e). (f) Posttranslocational glycosylation of extreme C-terminal sites occurs within a zone of glycosylation that is defined by the distance between the luminal surface of the ER membrane and the OST active site. Structural models for the STT3 (PglB) catalytic subunit of the OST complex and the Sec61 translocation channel (SecYE-β) are embedded in the membrane bilayer and scaled to the mean thickness of an ER cisternae. The PglB and SecYE-β structures were rendered with Mac PyMOL software using Protein Data Bank file 3RCE (*C. lari* PglB) and 1RHZ (*Methanocaldococcus jannaschii* SecYE-β). (g) Incomplete glycosylation of an extreme C-terminal site could be explained by exit of the nascent protein from the zone of glycosylation or folding of the polypeptide.

the direction of nascent polypeptide exit from the translocation channel, as the extreme C-terminal $N_{396}GT$ site is modified before the $N_{380}RS$ site. This observation excludes the possibility that glycosylation of C-terminal sites in proteins is achieved by reducing the rate of polypeptide egress from the lumenal face of the protein translocation channel after termination of protein synthesis. Thus, these sites are posttranslocationally glycosylated after the completed protein enters the ER lumen (Fig. 7 c). Second, our results indicate that the kinetics and extent of C-terminal glycosylation is strongly influenced by the hydroxyamino acid in the consensus site. Finally, the difference in modification kinetics of the two extreme C-terminal sites in SHBG is not consistent with a scanning mechanism wherein a single copy of STT3B engages a nascent polypeptide and scans for skipped NXT/S sites.

As the distance between a glycosylation site and the C terminus increases, the dependence upon STT3B decreased. The $N_{630}VT$ site in Tf, which is 68 residues from the C terminus, is efficiently glycosylated in STT3B-depleted cells. Analysis of several Tf derivatives and β-GUS Myc-DDK indicated that the boundary for STT3B dependence falls in the vicinity of 50–55 residues from the C terminus of the protein. Thus, hypoglycosylation of Tf by CDG-1 patients is explained by increased acceptor site skipping by STT3A in cells that lack the fully assembled oligosaccharide donor rather than a defect in posttranslocational glycosylation by STT3B. The previously analyzed posttranslational glycosylation site in factor VII (Bolt et al., 2005; Ruiz-Canada et al., 2009) is well outside the range for extreme C-terminal sites (84 residues from the C terminus).

Clearly, factors other than an extreme C-terminal location can cause glycosylation sites, such as the $N_{360}IT$ site factor VII to be skipped by STT3A.

What prevents complete modification of extreme C-terminal glycosylation sites, such as the $N_{380}RS$ site, in SHBG? For a slowly modified site, folding of the nascent polypeptide (Fig. 7 g) could prevent glycosylation as a result of the architecture of the OST active site (Lizak et al., 2011), which can only accommodate an unfolded region of a protein. A second factor that could limit posttranslocational glycosylation would be diffusion of the nascent polypeptide away from the lumenal surface of the ER membrane where the OST is located (Fig. 7 f). The distance between the lumenal surface of the ER membrane and the OST active site had been estimated to be 3 nm (Nilsson and von Heijne, 1993). However, this value is somewhat greater than the 1.5-nm distance between the acceptor peptide binding site and the transmembrane spans of *Campylobacter lari* PglB (Lizak et al., 2011). Electron microscopy of yeast cells indicates that the thickness of ER cisternae in the vicinity of membrane bound ribosomes is on average 36 nm (West et al., 2011). Thus, roughly 80% of the lumenal volume of the ER is outside a narrow zone of glycosylation near the lumenal membrane surface. C-terminal sites in glycoproteins that are tethered to the membrane by transmembrane spans or glycosylphosphatidylinositol anchors would be retained closer to the zone of glycosylation, likely explaining why the C-terminal sites in the two type II membrane proteins (CD40L and CD69) were less sensitive to STT3B depletion than the C-terminal sites in β-GUS or SHBG.

Internal glycans stimulate posttranslocational glycosylation by STT3B

A single internal glycan, regardless of location, was sufficient to increase the rate and extent of C-terminal glycosylation of the $N_{631}ET$ site in β-GUS. A similar enhancement of posttranslocational glycosylation by a second glycan was observed for two other substrates (factor VII and KCNE1) that we analyzed previously (Ruiz-Canada et al., 2009; Bas et al., 2011). A challenge for the future will be to determine whether posttranslocational delivery of substrates to STT3B is mediated by ER chaperones or instead simply occurs by diffusion (Fig. 7 c).

Conserved features of extreme C-terminal glycosylation

Analysis of the glycopeptide databases from seven model organisms revealed both conserved and nonconserved features of C-terminal glycosylation. As reported previously, there is strong positive selection for NXT sequons in glycoprotein sequences for organisms that have the ER–glycoprotein quality control pathway (Cui et al., 2009), and this includes all seven model organisms analyzed here. The reduction in C-terminal glycan density was most pronounced for the *M. musculus*, *D. rerio*, and *D. melanogaster* glycoproteins. The organisms that showed the least reduction in C-terminal glycan density (*S. cerevisiae*, *S. pombe*, and *C. elegans*) showed little or no reduction in sequon density between the C-terminal segment and internal regions of glycoproteins. Collectively with the phylogenetic analysis of STT3 proteins, it appears that posttranslocational glycosylation of sequons by an STT3B-like OST complex preceded the evolution of a dedicated translocation channel–associated STT3A complex.

Cellular roles of the STT3A and STT3B isoforms of the OST

The STT3A and STT3B isoforms of the OST complex have partially nonoverlapping roles in asparagine-linked glycosylation. The three internal glycosylation sites in β-GUS exemplify the class of sites that can be modified by either STT3A or STT3B, as hypoglycosylation of these internal sites only occurred when expression of both OST isoforms was reduced.

One prominent role of the STT3B complex is to mediate glycosylation of sequons that are skipped by the STT3A complex, including those at the extreme C terminus of proteins. For reasons that remain to be determined, the translocation channel–associated STT3A complex skips certain internal sequons. Previously characterized skipped sites include an N-terminal site in cathepsin C adjacent to the signal sequence cleavage site and a posttranslocational glycosylation site in blood coagulation factor VII (Ruiz-Canada et al., 2009). The internal $N_{111}AC$ site in CD69 is frequently skipped by STT3A, as NXC sites are inherently suboptimal because of a reduced binding affinity of the acceptor sequon to the OST active site (Bause, 1984). The $N_{111}AC$ site is also poorly modified by the STT3B complex, hence occupancy of this site remains low.

A novel role for the STT3B isoform of the OST in the ER-associated degradation pathway was recently described (Sato et al., 2012). Posttranslational glycosylation of a normally unmodified sequon in a folding-defective form of transthyretin directs the malfolded protein into the glycan-dependent branch of the ER-associated degradation pathway. It remains to be determined whether this aspect of the ER-associated degradation pathway is applicable to other malfolded proteins with silent glycosylation sites. These results, collectively with previous findings concerning a co-translational role for STT3A (Ruiz-Canada et al., 2009), indicate that the duplication of the *STT3* gene in metazoan organisms resulted in the evolution of OST isoforms with distinct roles in N-glycosylation that together facilitate optimal glycosylation of the large number of OST substrates that are expressed by higher eukaryotes.

Materials and methods

Cell culture and plasmid or siRNA transfection

HeLa cells (CCL-13; American Type Culture Collection) were cultured in 10-cm² dishes at 37°C in DMEM (Gibco), 10% fetal bovine serum with 100 U/ml penicillin and 100 µg/ml streptomycin. HeLa cells were seeded at 30% confluency for siRNA transfection or 80% confluency for plasmid transfection in 60-mm dishes and grown for 24 h before transfection with siRNA (60 nM negative control, 50 nM STT3A, and 60 nM STT3B) or 8 µg plasmid and Lipofectamine 2000 in Opti-MEM (Gibco) using a protocol from the manufacturer (Invitrogen). Plasmid transfection was performed after 48 h of siRNA transfection, and cells were assayed 24 h later. The siRNAs specific for the *STT3A* and *STT3B* mRNAs were characterized previously (Ruiz-Canada et al., 2009). The STT3A siRNAs are 5′-GGCCGUUUCUCUCA-CCGGCdTdT-3′ annealed with 5′-UCCGGUGAGAGAAACGGCCdTdT-3′. The STT3B siRNAs are 5′-GCUCUAUAUGCAAUCAGUAdTdT-3′ annealed with 5′-CACUGAUUUGCAUAUAGAGCdTdT-3′. Negative control siRNA was purchased from QIAGEN.

The Myc-DDK–tagged β-GUS, CD40 ligand, and CD69 expression vectors were purchased from OriGene. Tf was amplified from an OriGene cDNA clone and cloned into a pCMV6-AC-DDK-His vector (OriGene). The SHBG, N380Q, and N396Q SHBG mutants in the vector pRC/cytomegalovirus (Invitrogen) were gifts from G. Hammond (Child and Family Research Institute, Vancouver, British Columbia, Canada). Site-directed mutagenesis was used to generate glycosylation site mutants or to introduce a stop codon after the Myc or DDK tags in the expression vector. With the exception of the $N_{173}NT$ site in β-GUS that was inactivated by the T175A mutation, glycosylation sites were inactivated by asparagine to glutamine substitutions.

Radiolabeling and immunoprecipitation of glycoproteins

Cell culture medium was replaced with methionine- and cysteine-free DMEM media (Gibco) containing 10% dialyzed fetal bovine serum 20 min before the addition of 200 µCi/ml of Tran³⁵S label (PerkinElmer). Pulse-labeling periods were terminated by the addition of unlabeled methionine (3.75 mM) and cysteine (0.75 mM). Cells from one culture dish at each time point were lysed at 4°C by a 30-min incubation with 1 ml radioimmunoprecipitation assay lysis buffer or β-GUS lysis buffer (10 mM Tris-Cl, pH 8.5, 14 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 1 mM $MgCl_2$, and protease inhibitor cocktail, as defined in Kelleher et al. [1992]). Cell lysates were clarified by centrifugation (2 min at 13,000 rpm) and precleared by incubation for 2 h with control IgG and a mixture of protein A/G–Sepharose beads (Invitrogen). The precleared lysates were incubated overnight with protein or epitope tag–specific antibodies followed by the addition of a second aliquot of protein A/G–Sepharose beads and incubated for 4 h. Beads were washed five times with radioimmunoprecipitation lysis buffer or β-GUS wash buffer (150 mM Tris-Cl, pH 7.4, 500 mM NaCl, 0.4% NP-40, and 1% sodium deoxycholate) and twice with 10 mM Tris-HCl before eluting proteins with gel loading buffer. Antibodies were obtained from the following sources: anti-Tf (Dako), anti-DDK (anti-FLAG; Sigma-Aldrich), anti-SHBG (R&D Systems), and anti-Myc (Santa Cruz Biotechnology, Inc.). As indicated, immunoprecipitated proteins were digested with endoglycosidase H (New England Biolabs, Inc.). Dry gels were exposed to a phosphor screen (Fujifilm), scanned in a laser scanner (Typhoon FLA 9000; GE Healthcare), and quantified using AlphaEaseFC (Alpha Innotech).

Analysis of SHBG and β-GUS pulse–chase experiments

The pulse–chase data for β-GUS and SHBG glycosylation was fit to single and double exponential equations using Kaleidograph 3.6 (Synergy Software). For wild-type SHBG, which has two glycans, consumption of SHBG-0 was fit to the equation $y = a_0 \times e^{(-0.693t/t1/2(0))} + b_0$; formation of SHBG-2 was fit to the equation $y = a_2 \times (1 - e^{(-0.693t/t1/2(2))}) + b_2$. Consumption of SHBG-1 was fit to the equation $y = a_0 \times e^{(-0.693t/t1/2(0))} + a_2 \times (1 - e^{(-0.693t/t1/2(2))}) + b_1$ using the parameters derived for SHBG-0 and SHBG-2 (a_0, a_2, $t_{1/2(0)}$, and $t_{1/2(2)}$).

Bioinformatics analysis of glycoproteomic databases and STT3 proteins

A list of 5,052 high confidence (class I) murine N-glycosylation sites was obtained from the supplemental materials in Zielinska et al. (2010). We excluded 112 sites that did not match the currently accepted consensus sequon (N-X-T/S/C), as current evidence indicates that nonsequons, when modified, show occupancy levels that do not exceed 1–3% (Valliere-Douglass et al., 2009, 2010). 18 additional putative glycopeptides were excluded because they map to proteins that do not enter the secretory pathway (12 sites) or are no longer in the murine protein database as a result of a change in open reading frame annotation (six sites). The remaining 4,922 class I sites are in 1,902 murine proteins. The glycoprotein sequences were downloaded, and the locations of the glycopeptides were verified. When several nonidentical protein sequence files were obtained for a given glycopeptide, we selected the file with the largest open reading frame to avoid duplication (for alternatively spliced proteins) or analysis of incomplete protein sequences (N- or C-terminal truncations). The amino acid compositions of the 1,902 glycoproteins were used to calculate expected sequon densities. A complete list of 13,326 sequons (N-(X≠P)-T/S/C) in the 1,902 murine proteins was generated.

High confidence (class I) sites for *S. cerevisiae*, *S. pombe*, *C. elegans*, *D. melanogaster*, *A. thaliana*, and *D. rerio* were obtained from the supplemental materials in Zielinska et al. (2012) and analyzed in a similar manner. The generated sequon lists for all model organisms except *M. musculus* only included N-(X≠P)-T/S sequons because of the low number of N-X-C–linked glycans in the initial class I site databases of these model organisms.

Direct retrieval and BLAST (Basic Local Alignment Search Tool) search on the NCBI Protein and UniProt databases retrieved 999 eukaryotic STT3 sequences. Of these, 369 sequences were left after elimination of duplicates and severely truncated sequences. For display purposes (Fig. S1), the collection of 369 STT3 sequences was trimmed to 152 sequences representing diverse eukaryotes yet retaining multiple nonidentical STT3 sequences from selected metazoan, plant, and protists organisms. Phylogenetic alignment of the complete dataset (369 STT3 sequences) yielded a similar tree. The 152 sequences were used for phylogenetic analysis on the SATCHMO-JS web server (Hagopian et al., 2010). The tree drawing was performed using the ETE (python environment for phylogenetic tree exploration) program (Huerta-Cepas et al., 2010).

Online supplemental material

Fig. S1 shows the phylogenetic analysis of eukaryotic STT3 proteins. Table S1 lists the accession numbers of the 152 STT3 proteins that were used to generate Fig. S1. Online supplemental material is available at http://www.jcb.org/cgi/content/full/jcb.201301031/DC1.

References

Bas, T., G.Y. Gao, A. Lvov, K.D. Chandrasekhar, R. Gilmore, and W.R. Kobertz. 2011. Post-translational N-glycosylation of type I transmembrane KCNE1 peptides: implications for membrane protein biogenesis and disease. *J. Biol. Chem.* 286:28150–28159. http://dx.doi.org/10.1074/jbc.M111.235168

Bause, E. 1984. Model studies on N-glycosylation of proteins. *Biochem. Soc. Trans.* 12:514–517.

Ben-Dor, S., N. Esterman, E. Rubin, and N. Sharon. 2004. Biases and complex patterns in the residues flanking protein N-glycosylation sites. *Glycobiology.* 14:95–101. http://dx.doi.org/10.1093/glycob/cwh004

Bocchinfuso, W.P., K.L. Ma, W.M. Lee, S. Warmels-Rodenhiser, and G.L. Hammond. 1992. Selective removal of glycosylation sites from sex hormone-binding globulin by site-directed mutagenesis. *Endocrinology.* 131:2331–2336. http://dx.doi.org/10.1210/en.131.5.2331

Bolt, G., C. Kristensen, and T.D. Steenstrup. 2005. Posttranslational N-glycosylation takes place during the normal processing of human coagulation factor VII. *Glycobiology.* 15:541–547. http://dx.doi.org/10.1093/glycob/cwi032

Cui, J., T. Smith, P.W. Robbins, and J. Samuelson. 2009. Darwinian selection for sites of Asn-linked glycosylation in phylogenetically disparate eukaryotes and viruses. *Proc. Natl. Acad. Sci. USA.* 106:13421–13426. http://dx.doi.org/10.1073/pnas.0905818106

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. http://dx.doi.org/10.1016/j.molcel.2005.05.029

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. http://dx.doi.org/10.1093/protein/3.5.433

Haeuptle, M.A., and T. Hennet. 2009. Congenital disorders of glycosylation: an update on defects affecting the biosynthesis of dolichol-linked oligosaccharides. *Hum. Mutat.* 30:1628–1641. http://dx.doi.org/10.1002/humu.21126

Hagopian, R., J.R. Davidson, R.S. Datta, B. Samad, G.R. Jarvis, and K. Sjölander. 2010. SATCHMO-JS: a webserver for simultaneous protein multiple sequence alignment and phylogenetic tree construction. *Nucleic Acids Res.* 38(Suppl. 2):W29–W34. http://dx.doi.org/10.1093/nar/gkq298

Hershey, J.W.B. 1991. Translational control in mammalian cells. *Annu. Rev. Biochem.* 60:717–755. http://dx.doi.org/10.1146/annurev.bi.60.070191.003441

Huerta-Cepas, J., J. Dopazo, and T. Gabaldón. 2010. ETE: a python Environment for Tree Exploration. *BMC Bioinformatics.* 11:24. http://dx.doi.org/10.1186/1471-2105-11-24

Hülsmeier, A.J., P. Paesold-Burda, and T. Hennet. 2007. N-glycosylation site occupancy in serum glycoproteins using multiple reaction monitoring liquid chromatography-mass spectrometry. *Mol. Cell. Proteomics.* 6:2132–2138. http://dx.doi.org/10.1074/mcp.M700361-MCP200

Kelleher, D.J., and R. Gilmore. 2006. An evolving view of the eukaryotic oligosaccharyltransferase. *Glycobiology.* 16:47R–62R. http://dx.doi.org/10.1093/glycob/cwj066

Kelleher, D.J., G. Kreibich, and R. Gilmore. 1992. Oligosaccharyltransferase activity is associated with a protein complex composed of ribophorins I and II and a 48 kd protein. *Cell.* 69:55–65. http://dx.doi.org/10.1016/0092-8674(92)90118-V

Kelleher, D.J., D. Karaoglu, E.C. Mandon, and R. Gilmore. 2003. Oligosaccharyltransferase isoforms that contain different catalytic STT3 subunits have distinct enzymatic properties. *Mol. Cell.* 12:101–111. http://dx.doi.org/10.1016/S1097-2765(03)00243-0

Kowarik, M., S. Numao, M.F. Feldman, B.L. Schulz, N. Callewaert, E. Kiermaier, I. Catrein, and M. Aebi. 2006. N-linked glycosylation of folded proteins by the bacterial oligosaccharyltransferase. *Science.* 314:1148–1150. http://dx.doi.org/10.1126/science.1134351

Larkin, A., and B. Imperiali. 2011. The expanding horizons of asparagine-linked glycosylation. *Biochemistry.* 50:4411–4426. http://dx.doi.org/10.1021/bi200346n

Lizak, C., S. Gerber, S. Numao, M. Aebi, and K.P. Locher. 2011. X-ray structure of a bacterial oligosaccharyltransferase. *Nature.* 474:350–355. http://dx.doi.org/10.1038/nature10151

Matzuk, M.M., and I. Boime. 1988. The role of the asparagine-linked oligosaccharides in the secretion and assembly of human chorionic gonadotropin. *J. Cell Biol.* 106:1049–1059. http://dx.doi.org/10.1083/jcb.106.4.1049

McKinnon, T.A., E.C. Goode, G.M. Birdsey, A.A. Nowak, A.C. Chan, D.A. Lane, and M.A. Laffan. 2010. Specific N-linked glycosylation sites modulate synthesis and secretion of von Willebrand factor. *Blood.* 116:640–648. http://dx.doi.org/10.1182/blood-2010-02-267450

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. http://dx.doi.org/10.1083/jcb.200301043

Pult, F., G. Fallah, U. Braam, S. Detro-Dassen, C. Niculescu, B. Laube, and G. Schmalzing. 2011. Robust post-translocational N-glycosylation at the extreme C-terminus of membrane and secreted proteins in *Xenopus laevis* oocytes and HEK293 cells. *Glycobiology.* 21:1147–1160. http://dx.doi.org/10.1093/glycob/cwr013

Ruiz-Canada, C., D.J. Kelleher, and R. Gilmore. 2009. Cotranslational and post-translational N-glycosylation of polypeptides by distinct mammalian OST isoforms. *Cell*. 136:272–283. http://dx.doi.org/10.1016/j.cell.2008.11.047

Sato, T., Y. Sako, M. Sho, M. Momohara, M.A. Suico, T. Shuto, H. Nishitoh, T. Okiyoneda, K. Kokame, M. Kaneko, et al. 2012. STT3B-dependent posttranslational N-glycosylation as a surveillance system for secretory protein. *Mol. Cell*. 47:99–110.

Shibatani, T., L.L. David, A.L. McCormack, K. Frueh, and W.R. Skach. 2005. Proteomic analysis of mammalian oligosaccharyltransferase reveals multiple subcomplexes that contain Sec61, TRAP, and two potential new subunits. *Biochemistry*. 44:5982–5992. http://dx.doi.org/10.1021/bi047328f

Shipley, J.M., J.H. Grubb, and W.S. Sly. 1993. The role of glycosylation and phosphorylation in the expression of active human beta-glucuronidase. *J. Biol. Chem.* 268:12193–12198.

Tamura, T., J.H. Cormier, and D.N. Hebert. 2011. Characterization of early EDEM1 protein maturation events and their functional implications. *J. Biol. Chem.* 286:24906–24915. http://dx.doi.org/10.1074/jbc.M111.243998

Valliere-Douglass, J.F., P. Kodama, M. Mujacic, L.J. Brady, W. Wang, A. Wallace, B. Yan, P. Reddy, M.J. Treuheit, and A. Balland. 2009. Asparagine-linked oligosaccharides present on a non-consensus amino acid sequence in the CH1 domain of human antibodies. *J. Biol. Chem.* 284:32493–32506. http://dx.doi.org/10.1074/jbc.M109.014803

Valliere-Douglass, J.F., C.M. Eakin, A. Wallace, R.R. Ketchem, W. Wang, M.J. Treuheit, and A. Balland. 2010. Glutamine-linked and non-consensus asparagine-linked oligosaccharides present in human recombinant antibodies define novel protein glycosylation motifs. *J. Biol. Chem.* 285:16012–16022. http://dx.doi.org/10.1074/jbc.M109.096412

Vance, B.A., W. Wu, R.K. Ribaudo, D.M. Segal, and K.P. Kearse. 1997. Multiple dimeric forms of human CD69 result from differential addition of N-glycans to typical (Asn-X-Ser/Thr) and atypical (Asn-X-cys) glycosylation motifs. *J. Biol. Chem.* 272:23117–23122. http://dx.doi.org/10.1074/jbc.272.37.23117

Wacker, M., D. Linton, P.G. Hitchen, M. Nita-Lazar, S.M. Haslam, S.J. North, M. Panico, H.R. Morris, A. Dell, B.W. Wren, and M. Aebi. 2002. N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. *Science*. 298:1790–1793. http://dx.doi.org/10.1126/science.298.5599.1790

West, M., N. Zurek, A. Hoenger, and G.K. Voeltz. 2011. A 3D analysis of yeast ER structure reveals how ER domains are organized by membrane curvature. *J. Cell Biol.* 193:333–346. http://dx.doi.org/10.1083/jcb.201011039

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. http://dx.doi.org/10.1074/jbc.271.11.6241

Yan, Q., and W.J. Lennarz. 2002. Studies on the function of oligosaccharyl transferase subunits. Stt3p is directly involved in the glycosylation process. *J. Biol. Chem.* 277:47692–47700. http://dx.doi.org/10.1074/jbc.M208136200

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. http://dx.doi.org/10.1016/j.cell.2010.04.012

Zielinska, D.F., F. Gnad, K. Schropp, J.R. Wiśniewski, and M. Mann. 2012. Mapping N-glycosylation sites across seven evolutionarily distant species reveals a divergent substrate proteome despite a common core machinery. *Mol. Cell*. 46:542–548. http://dx.doi.org/10.1016/j.molcel.2012.04.031