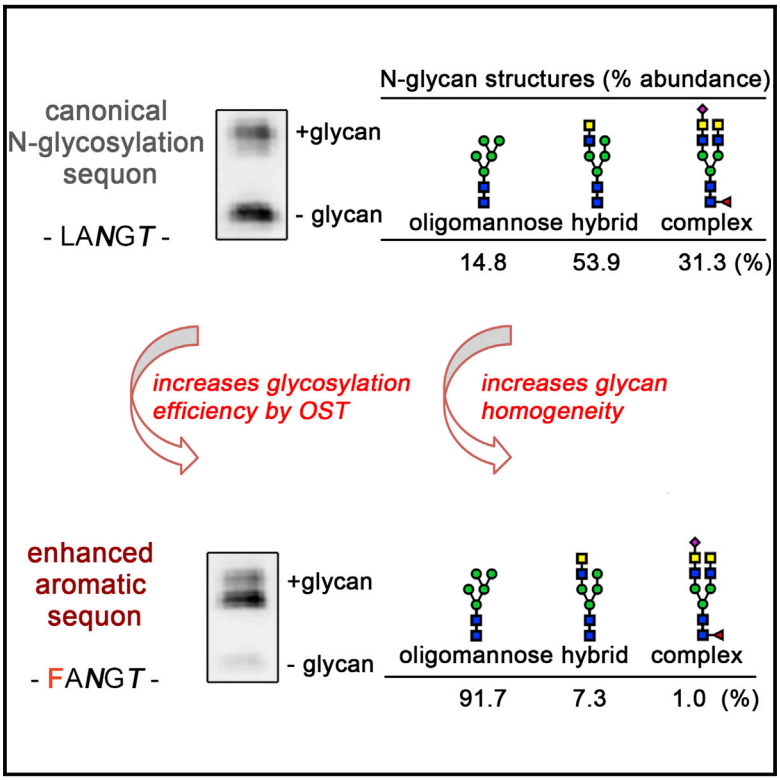
Article

Chemistry & Biology

Enhanced Aromatic Sequons Increase Oligosaccharyltransferase Glycosylation Efficiency and Glycan Homogeneity

Graphical Abstract



Highlights

- Aromatic amino acids two residues before sequons increase glycosylation efficiency
- Aromatic amino acids two residues before sequons decrease N-glycan heterogeneity
- Increased glycan occupancy results from oligosaccharyltransferase preferences
- Decreased N-glycan heterogeneity results from suppressed Golgi glycan remodeling

Authors

Amber N. Murray, Wentao Chen, Aristotelis Antonopoulos, ..., David L. Powers, Evan T. Powers, Jeffery W. Kelly

Correspondence

jkelly@scripps.edu (J.W.K.), epowers@scripps.edu (E.T.P.)

In Brief

Murray et al., using tandem N-glycoprotein repeats to eliminate intracellular processing effects, demonstrate that introducing an aromatic residue at *n*-2 relative to a canonical glycosylation sequon increases oligosaccharyltransferase N-glycosylation efficiency and suppresses Golgi glycan remodeling, affording more homogeneous N-glycans.

Murray et al., 2015, Chemistry & Biology *22*, 1–11
August 20, 2015 ©2015 Elsevier Ltd All rights reserved
http://dx.doi.org/10.1016/j.chembiol.2015.06.017

CellPress

Please cite this article in press as: Murray et al., Enhanced Aromatic Sequons Increase Oligosaccharyltransferase Glycosylation Efficiency and Glycan Homogeneity, Chemistry & Biology (2015), http://dx.doi.org/10.1016/j.chembiol.2015.06.017

Chemistry & Biology
Article

CellPress

Enhanced Aromatic Sequons Increase Oligosaccharyltransferase Glycosylation Efficiency and Glycan Homogeneity

Amber N. Murray,[1,2,6] Wentao Chen,[1,2,6] Aristotelis Antonopoulos,[3] Sarah R. Hanson,[1] R. Luke Wiseman,[2,4] Anne Dell,[3] Stuart M. Haslam,[3] David L. Powers,[5] Evan T. Powers,[1,*] and Jeffery W. Kelly[1,2,*]
[1]Department of Chemistry, The Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA
[2]Department of Molecular and Experimental Medicine, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA
[3]Department of Life Sciences, Imperial College London, London SW7 2AZ, UK
[4]Department of Chemical Physiology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA
[5]Department of Mathematics and Computer Science, Clarkson University, Potsdam, NY 13699, USA
[6]Co-first author
*Correspondence: jkelly@scripps.edu (J.W.K.), epowers@scripps.edu (E.T.P.)
http://dx.doi.org/10.1016/j.chembiol.2015.06.017

SUMMARY

N-Glycosylation plays an important role in protein folding and function. Previous studies demonstrate that a phenylalanine residue introduced at the *n*-2 position relative to an Asn-Xxx-Thr/Ser N-glycosylation sequon increases the glycan occupancy of the sequon in insect cells. Here, we show that any aromatic residue at *n*-2 increases glycan occupancy in human cells and that this effect is dependent upon oligosaccharyltransferase substrate preferences rather than differences in other cellular processing events such as degradation or trafficking. Moreover, aromatic residues at *n*-2 alter glycan processing in the Golgi, producing proteins with less complex N-glycan structures. These results demonstrate that manipulating the sequence space surrounding N-glycosylation sequons is useful both for controlling glycosylation efficiency, thus enhancing glycan occupancy, and for influencing the N-glycan structures produced.

INTRODUCTION

N-Glycosylation is a prevalent co- and/or post-translational modification of proteins that traverse the cellular secretory pathway (Apweiler et al., 1999). N-Glycans play crucial roles in protein homeostasis, modulating protein stability (Culyba et al., 2011; Hanson et al., 2009; Hebert et al., 2014; Imperiali and Rickert, 1995; Joao and Dwek, 1993; Price et al., 2012; Wang et al., 1996; Wormald and Dwek, 1999), chaperone-mediated folding (Hammond et al., 1994; Hebert et al., 2014; Helenius and Aebi, 2001; Jitsuhara et al., 2002; Oliver et al., 1997; Ou et al., 1993; Schmaltz et al., 2011; Ware et al., 1995), vesicular trafficking (Lu et al., 1997; Martina et al., 1998), and endoplasmic reticulum-associated degradation (ERAD) (Parodi, 2000; Vembar and Brodsky, 2008). N-Glycans also strongly influence biological function, including immune/inflammatory responses and circulating glycoprotein clearance, wherein carbohydrate binding proteins (lectins) recognize specific glycoforms of glycoproteins (Ahrens, 1993; Bertozzi and Kiessling, 2001; Chui et al., 2001; Sorensen et al., 2012).

During N-glycosylation, the heteromeric oligosaccharyltransferase enzyme complex (OST) in the lumen of the ER catalyzes the transfer of the $Glc_3Man_9GlcNAc_2$ oligosaccharide from a dolichol phosphate donor to the side-chain amide nitrogen of an acceptor Asn (Kornfeld and Kornfeld, 1985). Proteins translocated into the ER are scanned by OST for the N-glycosylation sequon Asn-Xxx-Ser/Thr (or, rarely, Asn-Xxx-Cys), where Xxx is any residue but Pro, and upon recognition of this sequon, the oligosaccharide is transferred to Asn either co- or post-translationally (Bas et al., 2011; Ruiz-Canada et al., 2009). Removal of two terminal glucoses from the N-glycan by ER glucosidases allows N-glycoproteins to engage the lectin chaperones calnexin and/or calreticulin, promoting proper folding (Caramelo and Parodi, 2008). Once folded, N-glycoproteins are packaged into vesicles and proceed to the Golgi, where they can interact with N-glycan processing enzymes that trim and elaborate glycans, affording a complex array of glycoforms (Lowe and Marth, 2003; Nairn et al., 2008).

It is estimated that one-third of sequons are not occupied by N-glycans (Apweiler et al., 1999; Petrescu et al., 2004; Surleac et al., 2012). It is not known whether signals exist in the structure of a protein that discourage OST from glycosylating proteins at a site that would render them unfoldable. Similarly, it is not known whether cues exist in a protein's primary or secondary structure that increase the glycosylation efficiency of OST for sequons in β turns, a context that can lead to an increase in protein stability (Culyba et al., 2011; Hanson et al., 2009; Price et al., 2011).

We have previously shown that introduction of a Phe residue at the *n*-2 position relative to the Asn of a sequon stabilizes glycosylated *Rattus norvegicus* cluster of differentiation 2 adhesion domain (CD2ad), *Homo sapiens* muscle acylphosphatase, and *Homo sapiens* Pin1 WW domain (Culyba et al., 2011; Hanson et al., 2009; Price et al., 2011). This stabilization is due to

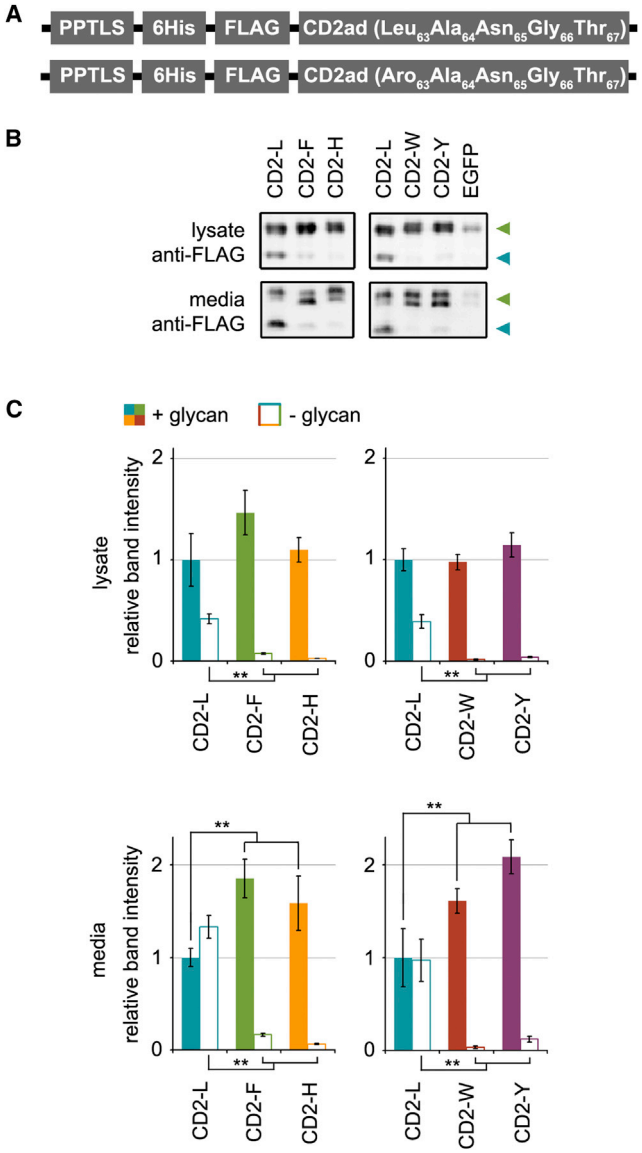


Figure 1. Aromatic Residues at *n*-2 Decrease the Amount of Non-glycosylated CD2ad and Increase the Amount of Secreted Glycosylated CD2ad

(A) CD2ad expression constructs, both CD2-L (Leu) and CD2-Aro (Aromatic) variants, allow for targeting to the secretory pathway via the PPTLS and western blot analysis via the FLAG tag.

(B) Anti-FLAG western blots of HEK 293 cell lysate and media display bands corresponding to glycosylated (green arrowheads) and non-glycosylated (blue arrowheads) CD2ad variants. A faint nonspecific band is detected by anti-FLAG in the cell lysate, as seen in the EGFP negative control lanes.

(C) Quantification of western blots shows that variants with EASs display less non-glycosylated protein in cell lysate and media compared with CD2-L. Band intensities are normalized to that of the CD2-L + glycan band. Solid bars quantify glycosylated protein and open bars quantify non-glycosylated protein. Error bars indicate SDs from three biological replicates; $^{**}p < 0.01$.

See also Figure S1.

protein-carbohydrate interactions driven largely by dispersion forces between Phe, the N-glycan, and Thr in the context of a reverse turn (Chen et al., 2013). These forces create an enhanced aromatic sequon (EAS) structural motif that can be incorporated into glycosylation-naïve proteins to confer glycosylation-dependent native-state stabilization (Culyba et al., 2011). Phe at *n*-2 also increases glycan occupancy in CD2ad and muscle acylphosphatase secreted from Sf9 insect cells (Culyba et al., 2011), although it is not known whether this results from decreased ERAD, increased folding and trafficking, or increased OST glycosylation efficiency. Statistical analysis of the primary sequence surrounding sequons indicates an increased occurrence of aromatic residues immediately preceding occupied sequons compared with unoccupied sequons (Petrescu et al., 2004; Surleac et al., 2012), suggesting a potential OST preference for aromatic residues at *n*-2. Moreover, the residue at *n*-2 must be Asp or Glu for bacterial OST to glycosylate a sequon, demonstrating that this position can influence glycosylation efficiency (Kowarik et al., 2006; Wacker et al., 2002).

Here, we report that human OST preferentially glycosylates substrates containing EASs. We express the rat ortholog of CD2ad, a protein that is natively glycosylation naïve, without or with an EAS containing Phe, His, Tyr, or Trp at *n*-2 in human cells, and observe increased glycan occupancy for all CD2ad EASs expressed in this human cell line. We also express a protein, transthyretin (TTR), with a native sequon that is poorly glycosylated, and observe an increase in glycan occupancy after introduction of Phe at *n*-2. To directly test the preference of OST for Phe at *n*-2, and to remove any contributions of cellular processing differences between differentially glycosylated proteins, we express tandem repeats of two different N-glycoproteins, CD2ad and fibroblast growth factor 9 (FGF9), each as a single polypeptide chain. These tandem repeat sequences show that, for both N-glycoproteins, the EAS is a better OST substrate than the standard sequon within the same polypeptide chain. Lastly, we report that EASs alter glycan processing, with different residues at *n*-2, leading to different glycoform product ensembles. Thus, our results show that the sequence surrounding N-glycosylation sequons affects both OST-mediated glycosylation and glycosyltransferase-mediated glycan elaboration in the Golgi, suggesting that primary sequence manipulation can be used to alter glycoprotein structure and function.

RESULTS

EASs Increase Glycan Occupancy of CD2ad

To probe the effect of aromatic residues at *n*-2 on glycan occupancy, we used the glycosylation-naïve rat ortholog of CD2ad as a model protein. Previously, we showed that an N-glycosylation sequon introduced into the type I β-bulge reverse turn of CD2ad is glycosylated in Sf9 insect cells, and we determined the folding energetics of its glycoforms (Culyba et al., 2011; Hanson et al., 2009). However, the effects of EASs on glycan occupancy in proteins expressed in human cells and the mechanisms involved are unknown. In HEK 293 cells, we expressed variants of CD2ad containing an N-terminal preprotrypsin leader sequence (PPTLS) to target the protein to the ER, an N-terminal hexahistidine tag preceding a FLAG tag, and an Asp67Thr mutation (to introduce an N-glycosylation sequon; Figure 1A). The wild-type sequence surrounding the glycosylation site is $Leu_{63}Ala_{64}Asn_{65}Gly_{66}Thr_{67}$, with Leu at *n*-2 (CD2-L). We created

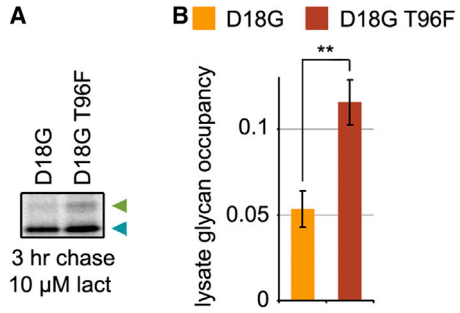


Figure 2. Phe at *n*-2 of D18G TTR Increases the Glycan Occupancy of the Poorly Glycosylated Native Sequon of TTR

(A) Autoradiogram of immunoprecipitated [^{35}S]D18G TTR variants from HEK 293 cell lysates displays bands corresponding to glycosylated (green arrowhead) and non-glycosylated (blue arrowhead) variants. Pulse-chase was performed in the presence of 10 μM lactacystin, and samples were harvested after a 3-hr chase.

(B) Quantification of bands shows that Phe at *n*-2 increases D18G TTR glycan occupancy. Error bars indicate SDs from three biological replicates; $^{**}p < 0.01$. See also Figure S2.

variants with Phe (CD2-F), His (CD2-H), Trp (CD2-W), or Tyr (CD2-Y) at *n*-2 to determine whether these aromatic residues increase glycan occupancy.

HEK 293 cells were transfected with expression vectors for CD2ad variants, and cell lysates and media were harvested 3 days after transfection. For each sample, cell lysate (25 μg of total protein) was subjected to SDS-PAGE and visualized by anti-FLAG western blotting. Cell media samples were analyzed similarly, where volumes loaded per lane were normalized to corresponding cell lysate protein concentrations (Figure 1B). The band intensities for glycosylated (green arrowheads) and non-glycosylated (blue arrowheads) CD2ad variants were quantified and normalized to the +glycan band of CD2-L (Figure 1C). In cell lysates, the +glycan band (solid bars) is not significantly affected by aromatic residues at *n*-2. However, the −glycan band (open bars) decreases for all aromatic variants. In media samples, there is significantly more +glycan band for all EAS-containing proteins and significantly less −glycan band for all EAS-containing proteins than for CD2-L. To ensure that glycosylated and non-glycosylated species were correctly identified, all variants were digested with PNGase F to remove N-glycans, causing the +glycan band intensities to decrease and the −glycan band intensities to increase (Figure S1). These results suggest that incorporating any aromatic residue at *n*-2 can increase N-glycan occupancy of proteins expressed in human cells.

EASs Increase Glycan Occupancy of D18G Transthyretin

We next attempted to boost the glycan occupancy of a poorly glycosylated native sequon by introducing Phe at *n*-2. Transthyretin (TTR) contains one native sequon in a type IV β turn, $Asn_{98}Asp_{99}Ser_{100}$, which is poorly glycosylated, owing both to Asp at *n*+1 and Ser at *n*+2 (Kasturi et al., 1997). D18G TTR is an amyloidogenic disease-associated mutant that is inefficiently secreted and subject to ERAD (Hammarstrom et al., 2003; Sekijima et al., 2005). This variant has been shown to be post-translationally glycosylated by OST to prevent aggregation and facilitate degradation (Sato et al., 2012).

We expressed FLAG-tagged variants of D18G TTR in HEK 293 cells, one with the wild-type sequence $Thr_{96}Ala_{97}Asn_{98}Asp_{99}Ser_{100}$ and one with Phe at *n*-2 (T96F). We then performed pulse-chase experiments in the presence of 10 μM lactacystin, to minimize proteasomal degradation effects. After a 20-min pulse and a 3-hr chase, cell lysates and media were harvested and [^{35}S]D18G TTR variants were immunoprecipitated with anti-FLAG affinity resin and quantified by phosphorimaging. D18G T96F glycan occupancy is twice that of D18G (0.12 ± 0.01 versus 0.06 ± 0.01) (Figures 2A and 2B). There is less total D18G than D18G T96F, indicating that proteasomal inhibition may be incomplete over the duration of the experiment or that D18G is being degraded in another way (Figure S2A). PNGase F digestion of samples causes the +glycan bands (green arrowhead) to disappear, indicating that this band corresponds to glycosylated TTR (Figure S2B). These results suggest that incorporating an EAS can increase glycan occupancy of a poor native sequon.

EASs Increase Glycan Occupancy in Single-Chain Tandem Repeats of CD2ad

The enhanced glycan occupancy of the EASs in CD2ad and TTR could come not only from glycosylation preferences of OST but also from differential intracellular processing. To determine more directly whether OST preferentially glycosylates EASs, we created single-chain tandem domain constructs that allow for expression of two CD2ad sequences within a single polypeptide chain. Because both CD2ad repeats are on the same polypeptide chain, they are unlikely to be differentially synthesized, trafficked, or degraded. Each construct contained a PPTLS, a FLAG tag, a CD2ad sequence, a tobacco etch virus (TEV) protease cleavage site, a hemagglutinin (HA) tag, and a second CD2ad sequence (Figure 3A). Quantitative comparison of the glycan occupancy of each CD2ad from the same polypeptide allowed us to determine the contribution of OST preference for particular sequons independent of other cellular processes. Four such constructs were created: L-F, where the N-terminal CD2ad has Leu at *n*-2 and the C-terminal CD2ad has Phe at *n*-2; F-L, where the N-terminal CD2ad has Phe at *n*-2 and the C-terminal CD2ad has Leu at *n*-2; L-L, where both domains have Leu at *n*-2; and F-F, where both domains have Phe at *n*-2.

At 3 days post-transfection in HEK 293 cells, cell lysates and media were harvested and analyzed by western blotting. When probed with either anti-FLAG antibody or anti-HA antibody, tandem proteins exhibit three bands, corresponding to species with two (red arrowheads), one (green arrowheads), or zero N-glycans (blue arrowheads) (Figure S3A). The disappearance of the higher two bands following PNGase F digestion, and the appearance of the lower band, confirms that the upper bands contain glycosylated tandem proteins (Figure S3B). Total intensity for all three bands was quantified for each sample to confirm that anti-FLAG (solid bars) and anti-HA (open bars) antibodies reported the same relative intensities (Figure S3C). In cell lysate samples, total tandem protein remains constant across constructs. However, total tandem protein in cell media samples differs, in that there is less L-L tandem protein and more F-F tandem protein (Figure S3C), which may reflect the known native-state stabilization provided by Phe at *n*-2 in CD2ad (Culyba et al., 2011).

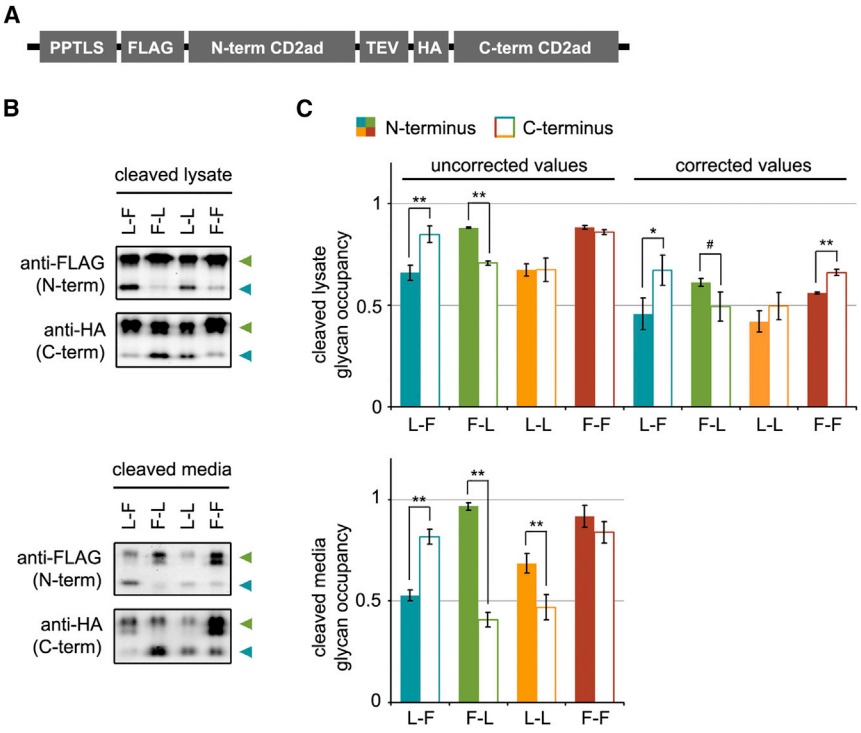


Figure 3. Single-Chain Tandem CD2ad Experiments Confirm Preference of OST for the EAS

(A) The tandem CD2ad cassette allows for expression of two CD2ad variants within a single polypeptide chain. Incubation with TEV protease cleaves the single-chain tandem protein into an N-terminal protein containing the FLAG epitope and the first CD2ad and a C-terminal protein containing the HA epitope and the second CD2ad.

(B) Anti-FLAG western blots of cell lysate and media display bands corresponding to glycosylated (green arrowheads) and non-glycosylated (blue arrowheads) N-terminal CD2ad. Anti-HA western blots display bands corresponding to glycosylated (green arrowheads) and non-glycosylated (blue arrowheads) C-terminal CD2ad.

(C) Quantification of western blots of cleaved CD2ad shows that glycan occupancy is higher for domains containing Phe at *n*-2 than for domains containing Leu at *n*-2, independent of whether the Phe-containing domain is N-terminal or C-terminal. Uncorrected values are glycan occupancies calculated from raw gel band intensities, and corrected values are corrected for low TEV cleavage efficiency of non-glycosylated tandem protein, as explained in the text. Solid bars quantify glycan occupancy of the N-terminal CD2ad and open bars quantify glycan occupancy of the C-terminal CD2ad. Error bars indicate SDs from three biological replicates; $^{\#}p = 0.05$, $^{*}p < 0.05$, $^{**}p < 0.01$. See also Figure S3 and Tables S1 and S2.

We next digested the tandem proteins in each cell lysate and each cell media sample using TEV protease. Cleaved samples were analyzed by western blot, where anti-FLAG reports on the N-terminal CD2ad and anti-HA reports on the C-terminal CD2ad (Figure 3B). Following TEV cleavage, the ~28-kDa bands corresponding to uncleaved tandem proteins decrease in intensity, and new ~14-kDa bands appear corresponding to cleaved proteins (Figure S3D). Glycan occupancy is higher for the CD2ad with Phe at *n*-2 than for the CD2ad with Leu at *n*-2, regardless of whether the Phe-containing CD2ad is at the N or C terminus in cell lysate or media samples (Figure 3C).

The decrease in intensity of uncleaved tandem proteins in lysate samples differs for the three ~28-kDa tandem protein bands, which correspond to doubly, singly, and non-glycosylated tandem proteins, indicating that the efficiency of TEV protease cleavage in lysate depends on the extent of glycosylation of the tandem proteins. TEV cleavage efficiency decreases with decreased glycosylation: doubly, singly, and non-glycosylated tandem proteins have cleavage efficiencies of ~70%–80%, ~40%–60%, and ~5%–15%, respectively (Table S1). We do not yet understand the cause of this differential cleavage efficiency, but because each pair of domains is part of a single polypeptide chain before TEV cleavage (Figure 3A), direct comparisons of glycan occupancies between the N- and C-terminal fragments following TEV cleavage are still meaningful. The uneven TEV protease cleavage efficiencies for a given tandem protein construct in lysate lead to over-representation of the absolute sequon occupancies for both the N- and C-terminal domains because of the relatively high cleavage efficiency of the doubly glycosylated tandem proteins and the relatively low cleavage efficiency of the non-glycosylated tandem proteins. However, the effect on the relative glycan occupancies of the N- and C-terminal domains is small. Furthermore, this direct comparison is independent of differences in total protein concentrations between different constructs (e.g., that more F-F is present than L-L), because only domains cleaved from a single polypeptide chain are analyzed.

We address the differential TEV cleavage efficiency in lysate in two ways. First, we have attempted to correct the gel band intensities of the 14-kDa glycosylated and non-glycosylated N- and C-terminal domains by subtracting the excess band intensity due to the high cleavage efficiency of the doubly glycosylated tandem protein from the glycosylated band, and by adding the deficient band intensity due to the low cleavage efficiency of the non-glycosylated tandem protein to the non-glycosylated band (Table S1). These corrections lower the observed glycan occupancy by ~0.2–0.3 in each case, but the differences between the sequon occupancies of domains with Leu at *n*-2 versus Phe at *n*-2 are maintained (Figure 3C, corrected values). Second, we have increased TEV cleavage efficiency by first immunoprecipitating tandem proteins from cell lysates using anti-FLAG affinity resin. The extent to which this increases the fraction of cleaved tandem protein is difficult to assess because of interference from the anti-FLAG light chain band, which is near the 28-kDa tandem protein bands (Figure S3E, asterisk). Nevertheless, quantification of the 14-kDa bands corroborates that the glycan occupancy is higher when Phe is the *n*-2 residue instead of Leu (compare Figure S3F with Figure 3C).

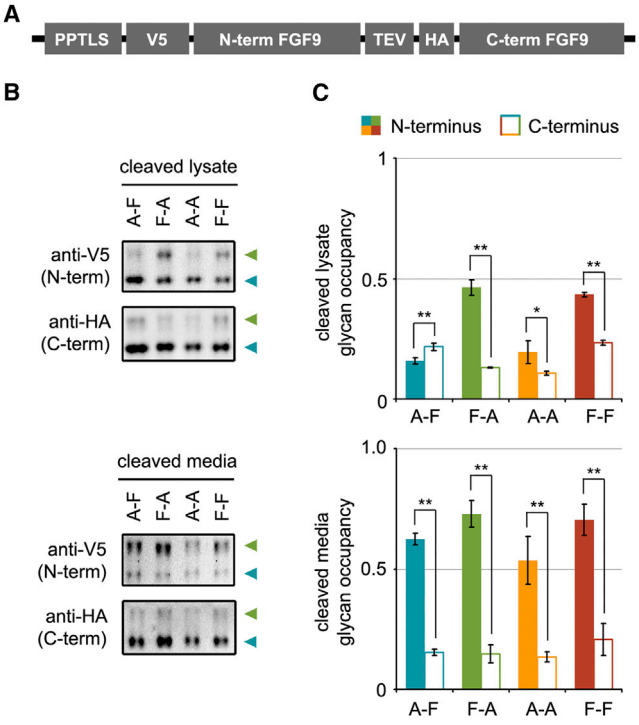


Figure 4. Single-Chain Tandem FGF9 Experiments Confirm Preference of OST for the EAS

(A) The tandem FGF9 cassette allows for expression of two FGF9 variants within a single polypeptide chain. Incubation with TEV protease cleaves the single-chain tandem protein into an N-terminal protein containing the V5 epitope and the first FGF9 and a C-terminal protein containing the HA epitope and the second FGF9.

(B) Anti-V5 western blots of cell lysate and media display bands corresponding to glycosylated (green arrowheads) and non-glycosylated (blue arrowheads) N-terminal FGF9. Anti-HA western blots display bands corresponding to glycosylated (green arrowheads) and non-glycosylated (blue arrowheads) C-terminal FGF9.

(C) Quantification of western blots of cleaved FGF9 shows that glycan occupancy is higher for domains containing Phe at *n*-2 than for domains containing Leu at *n*-2 from cell lysates, independent of whether the Phe-containing domain is N-terminal or C-terminal. Glycan occupancy in cell lysates is also higher in general for the N-terminal FGF9 than for the C-terminal FGF9, as highlighted in the A-A and F-F controls, independent of which residue is at *n*-2. In cell media, the glycan occupancy of the N-terminal domain overshadows EAS effects. Solid bars quantify glycan occupancy of the N-terminal FGF9 and open bars quantify glycan occupancy of the C-terminal FGF9. Error bars indicate SDs from three biological replicates; *$p < 0.05$, **$p < 0.01$.

See also Figure S4; Table S2.

EASs Increase Glycan Occupancy in Single-Chain Tandem Repeats of Fibroblast Growth Factor 9

To understand whether the OST preference for EASs is CD2ad specific or more universal, we created analogous single-chain tandem constructs using *Homo sapiens* fibroblast growth factor 9 (FGF9), a protein with a native EAS ($Phe_{77}Pro_{78}Asn_{79}Gly_{80}Thr_{81}$) within a type I β-bulge turn conformation, similar to that of CD2ad (Culyba et al., 2011) (Figure 4A). Initial experiments (data not shown) indicated that the Asn_{79}-Gly_{80}-Thr_{81} sequon for FGF9 domains exhibited nearly 100% glycan occupancy, irrespective of the presence or absence of Phe at the *n*-2 position. Therefore, the native Thr_{81} residue was mutated to Ser_{81}. Modifying the sequon to the less efficiently glycosylated Asn-Gly-Ser decreased "baseline" glycan occupancy (Kasturi et al., 1995; Petrescu et al., 2004) of the sequon to allow for a more sensitive measurement of the effects of the EAS on glycan occupancy. Four such constructs were created: A-F, where the N-terminal FGF9 has Ala at *n*-2 and the C-terminal FGF9 has Phe at *n*-2; F-A, where the N-terminal FGF9 has Phe at *n*-2 and the C-terminal FGF9 has Ala at *n*-2; A-A, where both domains have Ala at *n*-2; and F-F, where both domains have Phe at *n*-2. As is common practice when evaluating the effect of an amino acid side chain on a process of interest, Ala was chosen as the *n*-2 residue in the domains without an EAS to limit any side-chain contributions that might be introduced from other residues (Morrison and Weiss, 2001). For these tandem constructs, a V5 epitope precedes the N-terminal FGF9, and an HA epitope precedes the C-terminal FGF9.

At 2 days post-transfection in HEK 293T cells, cell lysates and media were harvested and analyzed by western blotting, as before. When probed with either anti-V5 antibody or anti-HA antibody, tandem proteins exhibit three bands, corresponding to species with two (red arrowheads), one (green arrowheads), or zero N-glycans (blue arrowheads) (Figure S4A). The disappearance of the upper two bands following PNGase F digestion, and the appearance of the lower band, confirms that the upper bands contain glycosylated tandem proteins (Figure S4B). The amount of tandem protein expressed between variants differs, with less overall protein for the A-A and F-F variants than for the A-F and F-A variants. However, total band intensity for all three bands is the same whether reported by anti-V5 (solid bars) or anti-HA (open bars) for each tandem protein (Figure S4C).

Following TEV cleavage, the ~40-kDa bands corresponding to uncleaved tandem proteins decrease in intensity, and new ~20-kDa bands appear corresponding to cleaved proteins (Figure S4D). The A-A and F-F controls indicate that the N-terminal FGF9 has higher glycan occupancy than the C-terminal FGF9 (Figure 4C), which is likely because sequons close to the C terminus are glycosylated relatively inefficiently (Ben-Dor et al., 2004; Gavel and von Heijne, 1990; Shrimal et al., 2013b). Sequons near the C terminus can be glycosylated post-translationally by only the STT3B-containing isoform of OST, whereas internal sequons can be glycosylated by both the STT3A- and STT3B-containing isoforms of OST (Shrimal et al., 2013b). Thus, the lower effective abundance of OST isoforms that are competent to glycosylate C-terminal sequons could explain their lower glycan occupancy. Nevertheless, in cell lysate samples, glycan occupancy is higher for the cleaved FGF9 with Phe at *n*-2 than for the cleaved FGF9 with Ala at *n*-2, regardless of whether the Phe-containing FGF9 domain is at the N or C terminus (Figures 4B and 4C). In cell media samples, the N-terminal glycosylation preference of OST overshadows EAS effects on sequon occupancy. These results reinforce our hypothesis that OST prefers EASs; however, they also suggest that EAS effects on glycan occupancy are context dependent.

To quantify the effect of EASs on glycan occupancy, independent of the effects of N- versus C-terminal domain placement, we fitted the glycan occupancy data to linear models with an indicator variable for the effect of Phe at *n*-2, an indicator variable for the effect of C-terminal placement of the EAS, and an

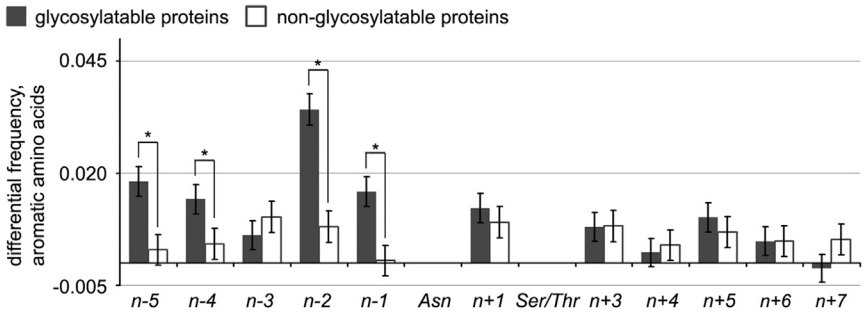


Figure 5. Aromatic Residues at *n*-2 Are Enriched in Glycosylatable Proteins Compared with Non-glycosylatable Proteins

The differential frequency of aromatic amino acids (frequency at given position minus overall frequency) for glycosylatable proteins that traverse the secretory pathway (solid bars) and non-glycosylatable proteins in the cytosol/mitochondria (open bars) are plotted versus their position relative to the Asn of N-glycosylation sequons. Error bars indicate SD; *p < 0.01 for a large sample, one-tailed test of the differences between binomial proportions.

See also Figure S5.

interaction term, to determine whether these effects are additive or whether they modulate each other (see Supplemental Information for details). These models fit the data well and indicate that Phe at *n*-2 significantly increases glycan occupancy for FGF9 lysate, FGF9 media, CD2ad lysate, and CD2ad media. These models also indicate that the effect of the EAS is attenuated for FGF9 when the EAS is present in the C-terminal domain. This effect is not apparent in CD2ad. The results of the regression are summarized in Table S2.

EASs Are Selected for in Proteins that Can Be N-Glycosylated

Given that OST glycosylation efficiency is increased in EASs, a selection pressure may exist that favors aromatic amino acids at *n*-2 in proteins that can be N-glycosylated (i.e., proteins that traverse the secretory pathway). To evaluate this hypothesis, we compared the frequencies of amino acids in positions five residues immediately before and after Asn-Xxx-Ser/Thr sequences between proteins that can and cannot be N-glycosylated (“glycosylatable” versus “non-glycosylatable” proteins). We defined glycosylatable proteins as those that have a gene ontology (GO) annotation indicating that they reside in the ER, the Golgi, the lysosome, the plasma membrane, or the extracellular space. We defined non-glycosylatable proteins as those that have a GO annotation indicating that they reside in the cytosol or the mitochondria. We excluded proteins present in both lists.

The final set of glycosylatable proteins comprises 3,076 proteins, with a mean length of 562 amino acids. The sequences of these proteins include 9,770 Asn-Xxx-Ser/Thr sequences that are more than five residues away from the N and C termini, for an overall occurrence rate of 0.62 sequons per 100 amino acids. The final list of non-glycosylatable proteins consists of 3,198 proteins, with a mean length of 575 amino acids. These sequences include 7,766 Asn-Xxx-Ser/Thr sequences that are more than five residues away from the N and C termini, for an overall occurrence rate of 0.45 Asn-Xxx-Ser/Thr sequences per 100 amino acids.

The overall frequencies of aromatic amino acids (Phe, Tyr, Trp, His) at all positions within the proteome are 0.109 and 0.097 for glycosylatable and non-glycosylatable proteins, respectively. The difference between the overall frequency of aromatic amino acids and the frequency of aromatic amino acids at positions near Asn-Xxx-Ser/Thr is shown in Figure 5 (raw frequency data are shown in Figure S5.) For glycosylatable proteins, aromatic amino acids are enriched generally in positions N-terminal to potential sequons. The p values that result from testing the hypothesis that aromatic amino acids are more enriched at positions *n*-5 to *n*-1 in glycosylatable than in non-glycosylatable proteins are <0.01 at all positions except for *n*-3, where there is no discernible enrichment of aromatic amino acids (Figure 5). This effect is especially strong at *n*-2, where the enrichment of aromatic amino acids is approximately 4-fold higher in glycosylatable than in non-glycosylatable proteins (0.0346 ± 0.0036 versus 0.0081 ± 0.0035). This result shows that, on a genome-wide scale, aromatic amino acids are favored at the *n*-2 position relative to Asn-Xxx-Ser/Thr sequences in glycosylatable proteins, consistent with the hypothesis that EASs are selected for to improve OST glycosylation efficiency.

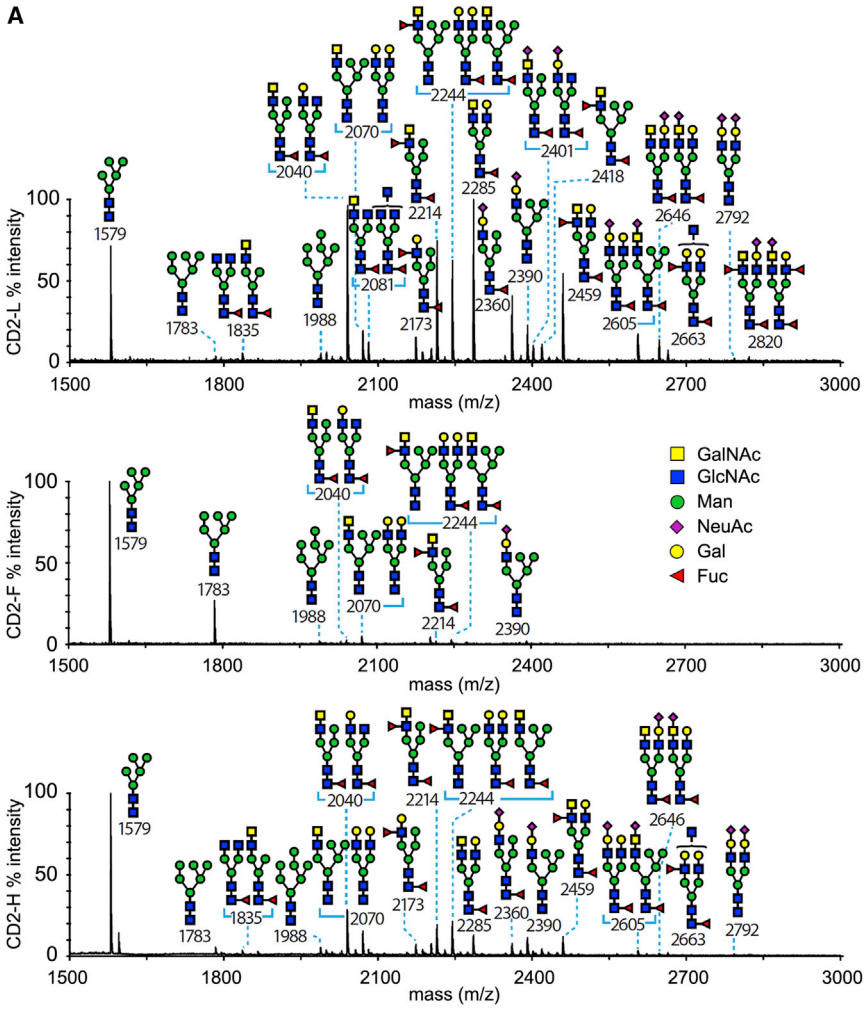
EASs Affect the N-Glycan Structures in CD2ad and FGF9

In addition to affecting OST glycosylation efficiency, aromatic amino acids at *n*-2 alter the SDS-PAGE banding patterns of glycosylated CD2ad harvested from cell media. Whereas glycosylated, secreted CD2-L exhibits predominantly one band, aromatic variants exhibit two bands (Figure 1B; Figure S6A), which collapse to the non-glycosylated size after digestion with PNGase F (Figure S6B). We hypothesized that residues at *n*-2 might influence the activities of Golgi enzymes involved in N-glycan processing, leading to different glycoform populations that have different gel banding patterns.

We analyzed the glycan composition of CD2-L, CD2-F, and CD2-H purified from HEK 293 media (Figure S6) using MALDI-TOF-mass spectrometry (MS) glycoform analysis (Ceroni et al., 2008; Jang-Lee et al., 2006). A multi-institutional assessment of glycomics methodologies demonstrated that MALDI-TOF-MS is a reliable method for relative quantification based on the signal intensities of permethylated glycans, especially when comparing signals over a small mass range within the same spectrum (Wada et al., 2007). Therefore, while absolute quantification of candidate N-glycans by this method is not strictly possible, valuable quantitative information can be extracted from MS data.

MALDI-TOF-MS demonstrated that CD2-L has a diverse array of glycoforms, including 14.8% oligomannose, 53.9% hybrid, and 31.3% complex N-glycans (Figures 6A [top panel] and 6B). In the EAS-containing CD2-F variant, glycoforms are far more homogeneous, comprising 91.7% oligomannose N-glycans (Figures 6A [middle panel] and 6B). This observation is consistent with previous work on the EAS-containing human ortholog of CD2ad, which has been shown to retain the oligomannose glycoforms when expressed in Chinese hamster ovary cells (Recny et al., 1992; Rudd et al., 1999; Wyss et al., 1995). CD2-H has a

A



B

	N-glycan structures (% abundance)		
	oligomannose	hybrid	complex
CD2-L	14.8	53.9	31.3
CD2-F	91.7	7.3	1.0
CD2-H	46.1	39.0	14.9

Figure 6. Different Residues at *n*-2 Alter Glycoform Populations of Secreted Variants of CD2ad

(A) MALDI-TOF mass spectra of permethylated N-glycans released by PNGase F digestion indicate that CD2-L N-glycans are highly heterogeneous and comprise a diverse set of oligomannose, hybrid, and complex structures. The heterogeneity is greatly reduced for CD2-F N-glycans, which are mainly oligomannose. The heterogeneity of CD2-H N-glycans is intermediate between that of CD2-L and that of CD2-F. All molecular ions are $[M + Na]^+$. Putative structures are based on composition, tandem MS, and biosynthetic knowledge. Structures that show sugars within a bracket have not been unequivocally defined.

(B) Comparison of N-glycan peak intensities confirms the reduction in complex glycoforms for CD2-F and CD2-H variants, compared with CD2-L. See also Figure S6.

glycoform population of intermediate complexity, with 46.1% oligomannose, 39.0% hybrid, and 14.9% complex N-glycans (Figures 6A [bottom panel] and 6B).

Given that the MALDI-TOF-MS results indicate that CD2-F is more homogeneous than CD2-L, it is perhaps surprising that CD2-F runs as two bands on SDS-PAGE while CD2-L runs as a single band. To further understand these observations, we analyzed the gel-extracted upper and lower +glycan CD2-F gel bands by MALDI-TOF-MS. We found that the upper band contains multiple glycoforms (including oligomannose), whereas the lower band contains only oligomannose glycans (Figure S6C). Thus, it appears that the oligomannose glycoforms of our CD2 constructs co-migrate with the more processed hybrid and complex-type glycoforms up to a point, but when the oligomannose glycoforms are highly enriched, as they are in CD2-F, the excess runs as a separate band. To corroborate this notion, we treated purified samples of CD2-F, CD2-L, and CD2-H with α1-2,3-mannosidase (which catalyzes the hydrolysis of α1-2 and α1-3 linked mannose residues from N-glycans). This treatment changed the gel mobility of the lower CD2-F +glycan band, while the mobilities of all of the other CD2ad variant bands were insensitive to mannosidase cleavage (Figure S6B). While it is clear that the digestion of the N-glycans with mannosidase was not complete (since the mobilities of the oligomannose glycoforms in the upper bands did not change), this result nevertheless qualitatively corroborates the MS findings that the oligomannose glycoforms are highly enriched in CD2-F.

To test whether EAS alteration of glycoform populations extends to other proteins, we used MALDI-TOF-MS to analyze the glycan composition of FGF9-A (with Ala at *n*-2) and FGF9-F (with Phe at *n*-2) secreted from HEK 293F cells. FGF9-A and FGF9-F were expressed as fusions to non-glycosylated immunoglobulin G Fc to allow affinity purification using Protein A (Figure S7). FGF9-A has a more diverse array of glycoforms, including 6.0% oligomannose, 2.0% hybrid, and 92.0% complex N-glycans, whereas the EAS-containing FGF9-F variant has less glycoform heterogeneity, with 1.3% oligomannose, 20.9% hybrid, and 77.8% complex N-glycans (Figures 7A and 7B).

Thus, in both CD2ad and FGF9, EASs alter Golgi elaboration of N-glycans, shifting glycoform populations away from

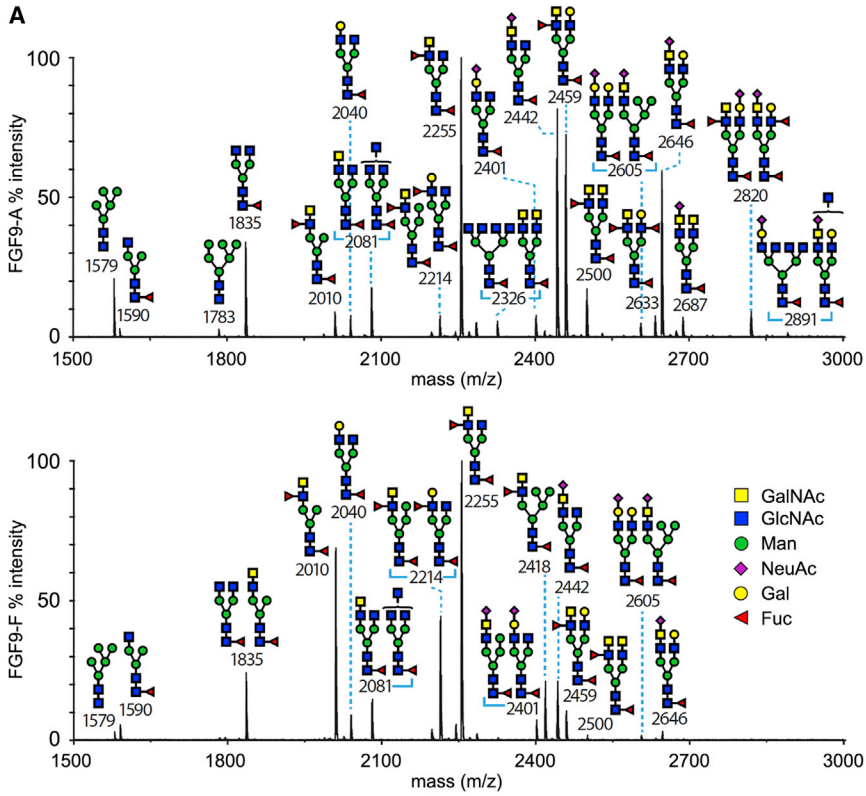


Figure 7. Different Residues at *n*-2 Alter Glycoform Populations of Secreted Variants of FGF9-Fc Fusions

(A) MALDI-TOF mass spectra of permethylated N-glycans released by PNGase F digestion indicate that FGF9-A N-glycans are more complex than FGF9-F N-glycans. All molecular ions are $[M + Na]^+$. Putative structures are based on composition, tandem MS, and biosynthetic knowledge. Structures that show sugars within a bracket have not been unequivocally defined.

(B) Comparison of N-glycan peak intensities confirms the reduction in complex glycoforms for FGF9-F compared with FGF9-A.

See also Figure S7.

B

	N-glycan structures (% abundance)		
	oligomannose	hybrid	complex
FGF9-A	6.0	2.0	92.0
FGF9-F	1.3	20.9	77.8

complex N-glycans and toward hybrid or oligomannose N-glycans. These results indicate that the primary sequence surrounding an N-glycosylation sequon modulates which glycoforms are produced.

DISCUSSION

N-Glycosylation greatly expands the functionality of the proteome. The human N-glycome is much more complex than the human proteome (Bertozzi and Kiessling, 2001; Freeze, 2006), and differences in glycan occupancy and glycoform diversity exist between copies of the same protein produced by a given cell. A vast N-glycan processing network containing more than 100 enzymes affects both glycosylation efficiency and subsequent N-glycan processing. Deficiencies in this network account for many congenital disorders of glycosylation (Chui et al., 1997; Jaeken, 2010; Shrimal et al., 2013a) and loss-of-function diseases (Glozman et al., 2009), while alterations can generate new biological function. Thus, variability in glycan occupancy and diversity strongly influence biological function.

We have shown that the primary sequence surrounding an N-glycosylation sequon strongly affects how efficiently it is glycosylated by OST. EASs are preferred by OST, exhibit higher glycan occupancy, and can result in native-state stabilization (Culyba et al., 2011; Hanson et al., 2009; Price et al., 2011). OST may therefore have evolved to prefer EASs, because of the likelihood that glycosylating such proteins would stabilize the native state and increase the overall levels of functional protein.

We have also shown that the primary sequence surrounding a sequon in a folded protein can strongly affect N-glycan processing. The extent to which an N-glycan is processed is determined, at least in part, by the accessibility of the N-glycan to the Golgi-resident glycan elaboration and/or trimming enzymes (Hsieh et al., 1983; Hubbard, 1988; Rudd et al., 1999; Thaysen-Andersen and Packer, 2012; Trimble et al., 1983). Based on the decrease in complex-type glycoforms exhibited in EAS-containing variants of CD2ad and FGF9, it appears that an interaction between an N-glycan and a single residue side chain, like the interaction between the N-glycan and an aromatic side chain in an EAS (Chen et al., 2013; Culyba et al., 2011; Price et al., 2011), can decrease the dynamic accessibility of an N-glycan enough to profoundly alter the extent to which it is processed in the Golgi.

Since EASs are excellent substrates for OST-mediated glycosylation and can reduce glycan elaboration in the secretory pathway, EASs can influence both intrinsic protein stabilization and the N-glycan structures produced. EASs can be engineered into therapeutic proteins to increase stability, glycan occupancy, and glycoform homogeneity, boosting total protein yields and decreasing immunogenicity. In addition, manipulating the sequence space surrounding N-glycosylation sequons in therapeutic proteins may provide control over half-life and clearance by shifting glycoform populations toward or away from structures that can be cleared by lectins (Sorensen et al., 2012). There are likely many other primary and secondary protein

structure effectors of the glycoproteome that are yet to be discovered. Further exploration of the EAS, and perhaps the flanking sequence, should elucidate engineering guidelines for enhancing glycosylation by OST and provide strategies for controlling glycan structural homogeneity. This structure-based approach offers the exciting prospect of manipulating protein stability, immunogenicity, targeting, and a host of other properties conveyed by protein N-glycosylation.

SIGNIFICANCE

N-Glycosylation, the OST-mediated conjugation of oligosaccharides onto the Asn side-chain NH of proteins, occurs in the majority of proteins that pass through the cellular secretory pathway. These N-glycans add an additional level of functional and structural diversity to the eukaryotic proteome. N-Glycans affect protein folding intrinsically through native-state protein-glycan interactions and extrinsically by influencing N-glycoprotein engagement with proteostasis network pathways, including chaperone-mediated folding and proteasomal degradation pathways. N-Glycans also engage with glycan binding proteins, affecting myriad biological processes. Herein we report that incorporating an aromatic amino acid (Phe, Tyr, His, or Trp) at the *n*-2 position relative to an N-glycosylation sequon, thereby creating an EAS, increases sequon occupancy, due to OST substrate preferences rather than differences in other cellular processing events such as degradation or trafficking. Moreover, the preference for aromatic amino acids at the *n*-2 position is evident not only biochemically but also evolutionarily, given the increased frequency of aromatic amino acids in this position in glycosylatable versus non-glycosylatable proteins. The EAS also influences the composition of the attached N-glycans by altering glycan remodeling in the Golgi. The tandem domain experimental approach employed is also significant because it should be useful for further exploration of the sequence preferences of OST and those of the Golgi N-glycan processing enzymes. In summary, this approach offers the opportunity to manipulate N-glycoprotein structure, stability, immunogenicity, and targeting, thereby augmenting protein therapeutic stability, shelf life, and half-life via sequence engineering.

EXPERIMENTAL PROCEDURES

Expression Construct Preparation

Standard protocols were used for vector construction. In brief, all expression constructs were cloned into vectors allowing for cytomegalovirus promoter-driven transcription. Mutations were introduced via QuikChange site-directed mutagenesis (Agilent Technologies 200515). The protocols for vector construction are described in detail in Supplemental Experimental Procedures.

Cell Culture

CD2ad-containing proteins were produced from HEK 293 cells transfected with FuGENE 6 (Roche 11814443001) and cultured as monolayers in D-MEM/F12 (Gibco 11330032) supplemented with 1% penicillin-streptomycin-glutamine (Gibco 10378-016) and 10% fetal bovine serum (Omega Scientific FB-02) media under standard conditions. Tandem FGF9 repeat proteins were produced from HEK 293T cells transfected with Xtreme Gene 9 (Roche 06365787001) and cultured as described above. FGF9-Fc fusions were produced from suspension HEK 293F cells transfected with 293Fectin (Invitrogen 12347-019) and cultured in FreeStyle 293 (Invitrogen 12338) under standard conditions. A detailed protocol for the pulse-chase experiments is available in Supplemental Experimental Procedures.

Purification of CD2ad and FGF9 Variants

CD2-L, CD2-F, and CD2-H were expressed as described above. Media from transfected cells were harvested every 2 days, and cells were split and expanded as necessary to maintain 50%–80% confluency. Total harvested media volume for each variant was ~500 ml. Urea powder was added to media samples to 8 M to denature proteins and reduce interactions between CD2ad variants and BSA or other secreted proteins. Variants were purified from denatured media samples by fast protein liquid chromatography. Samples were first run over 5-ml Histrap FF columns (GE Healthcare 17-5255-01) in 8 M urea, 150 mM NaCl, and 30 mM Bis-Tris (pH 6.8). CD2ad was eluted with a linear gradient of imidazole. Fractions containing CD2ad were pooled and concentrated in Amicon Ultra-15 Centrifugal Filter Units (Millipore UFC901008). 1 mM tris-(2-carboxyethyl)phosphine (TCEP) was added to concentrated samples, which were then purified by size-exclusion chromatography in 8 M urea, 30 mM Tris (pH 7.4) and 0.5% SDS using a Superdex 200 column (GE Healthcare 17-5175-01). Fractions containing CD2ad were again concentrated, dialyzed against 50 mM ammonium bicarbonate (pH 8.4), and lyophilized. Samples were analyzed for purity by SDS-PAGE and silver staining (Pierce 21642).

FGF9-A and FGF9-F were expressed in HEK 293F suspension cultures, harvested 6 days after transfection, and centrifuged at 6,000 × *g* for 30 min to pellet cells before filtering by 0.22-μm filter to remove residual cells. Supernatant was run over a 5-ml HiTrap Protein A column (GE 17-0403-01), and Fc-fusion proteins were eluted with 0.1 M Glycine (pH 2.7). Following dialysis into PBS, purified Fc-fusion proteins were analyzed for purity by SDS-PAGE and Coomassie blue staining.

Western Blotting

Three days post-transfection for CD2ad proteins and 2 days post-transfection for FGF9 proteins, media were harvested and centrifuged at 1,000 × *g* for 2 min to pellet cell contaminants, after which the supernatant was transferred to fresh tubes. Cells were harvested at the same time from culture dishes by trypsinization (Gibco 12605-010) and centrifuged at 1,000 × *g* for 2 min. Cell pellets were washed with Dulbecco's PBS (DPBS) (Gibco 14190-136) and lysed in lysis buffer (DPBS, 1% Triton X-100, 1 mM EDTA) for 30 min at 4°C. Lysed cells were centrifuged at 10,000 × *g* for 10 min, and post-nuclear supernatants were transferred to fresh tubes. Total protein in cell lysates was quantified by Micro BCA (Thermo Scientific 23235). For PNGase F digests, samples were incubated with 1 μl of PNGase F (NEB P0704S) for 2 hr at 37°C, according to the manufacturer's instructions. For TEV digests, samples were incubated with 2 μl of AcTEV Protease (Invitrogen 12575-015) at 30°C overnight, according to the manufacturer's instructions. For mannosidase digests, purified CD2ad variants (25 ng) were incubated with 0.5 μl of α1-2,3 mannosidase (NEB P0729S) for 2 hr at 37°C, according to the manufacturer's instructions. For CD2ad single-domain constructs and CD2ad and FGF9 tandem constructs, Laemmli buffer containing 1 mM TCEP was added to 25 μg of lysate per sample, and media sample volumes were proportional to the total protein concentrations of their respective lysates. For tandem constructs, Laemmli buffer containing 1 mM TCEP was added to 10 μg of lysate per sample, and media sample volumes were proportional to the amount of cells present in their respective lysates. For purified CD2ad variants, Laemmli buffer containing 1 mM TCEP was added to 25 ng of purified protein. Lanes labeled "C" contain 20 ng of non-glycosylated CD2ad control expressed in and purified from *Escherichia coli* (Culyba et al., 2011). All samples were boiled in Laemmli buffer for 5 min and subjected to SDS-PAGE using 4%–12% or 12% NuPAGE Novex Bis-Tris pre-cast gels (Invitrogen NP0323BOX, NP0343BOX) and MOPS running buffer (Invitrogen NP0001). Proteins were transferred to nitrocellulose membranes, which were stained with Ponceau S to visualize sample loading and transfer. Western blotting then proceeded using Odyssey blocking buffer (LI-COR 927-40000) and mouse anti-FLAG antibody (Agilent 200472-21, 1:2000),

mouse anti-HA antibody (Covance MMS-101P, 1:2000), or mouse anti-V5 antibody (Life Technologies R960-25, 1:5000). Proteins were visualized and quantified using goat anti-mouse IRDye 800CW (LI-COR 827-08364, 1:15,000) and the Odyssey Infrared Imaging System.

Glycomic Analysis

Glycomic analysis was carried out using GlycoWorkBench according to previous protocols (Ceroni et al., 2008; Jang-Lee et al., 2006). A detailed protocol for the glycomic analysis is available in Supplemental Experimental Procedures.

Linear Regression Models and Amino Acid Frequency Analysis

Detailed protocols for the linear regression modeling of glycan occupancies and amino acid frequency analysis are available in Supplemental Experimental Procedures.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.chembiol.2015.06.017.

AUTHOR CONTRIBUTIONS

A.N.M. and W.C. designed and executed experiments testing glycan occupancy/glycosylation efficiency, and expressed and purified proteins for glycoform analysis. A.A. and S.M.H. identified all glycoforms described. S.R.H. and R.L.W. helped to design experiments. D.L.P. performed statistical analyses of the data. E.T.P. performed statistical analyses of the data and amino acid frequency analysis, and helped to design experiments. A.D. and J.W.K. directed the project. A.N.M., E.T.P., and J.W.K. wrote the manuscript. All authors edited the manuscript.

ACKNOWLEDGMENTS

The authors thank Professor Ian A. Wilson for the Fc-fusion expression vector, Dr. Joseph Genereux for the D18G TTR expression construct, Dr. Elizabeth Culyba for purified non-glycosylated CD2ad, and Dr. Colleen Fearns for critical reading of and assistance in the preparation of this publication. This work was supported by the NIH R01 grant GM051105 (to J.W.K. and E.T.P), the Skaggs Institute for Chemical Biology, the Lita Annenberg Hazen Foundation, and the Biotechnology and Biological Sciences Research Council grant BB/K016164/1 (to A.D. and S.M.H).

Received: January 21, 2015
Revised: June 9, 2015
Accepted: June 11, 2015
Published: July 16, 2015

REFERENCES

Ahrens, P.B. (1993). Role of target cell glycoproteins in sensitivity to natural killer cell lysis. J. Biol. Chem. *268*, 385–391.

Apweiler, R., Hermjakob, H., and Sharon, N. (1999). On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. Biochim. Biophys. Acta *1473*, 4–8.

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

Ben-Dor, S., Esterman, N., Rubin, E., and Sharon, N. (2004). Biases and complex patterns in the residues flanking protein N-glycosylation sites. Glycobiology *14*, 95–101.

Bertozzi, C.R., and Kiessling, L.L. (2001). Chemical glycobiology. Science *291*, 2357–2364.

Caramelo, J.J., and Parodi, A.J. (2008). Getting in and out from calnexin/calreticulin cycles. J. Biol. Chem. *283*, 10221–10225.

Ceroni, A., Maass, K., Geyer, H., Geyer, R., Dell, A., and Haslam, S.M. (2008). GlycoWorkbench: a tool for the computer-assisted annotation of mass spectra of glycans. J. Proteome Res. *7*, 1650–1659.

Chen, W., Enck, S., Price, J.L., Powers, D.L., Powers, E.T., Wong, C.H., Dyson, H.J., and Kelly, J.W. (2013). Structural and energetic basis of carbohydrate-aromatic packing interactions in proteins. J. Am. Chem. Soc. *135*, 9877–9884.

Chui, D., Oh-Eda, M., Liao, Y.F., Panneerselvam, K., Lal, A., Marek, K.W., Freeze, H.H., Moremen, K.W., Fukuda, M.N., and Marth, J.D. (1997). Alpha-mannosidase-II deficiency results in dyserythropoiesis and unveils an alternate pathway in oligosaccharide biosynthesis. Cell *90*, 157–167.

Chui, D., Sellakumar, G., Green, R., Sutton-Smith, M., McQuistan, T., Marek, K., Morris, H., Dell, A., and Marth, J. (2001). Genetic remodeling of protein glycosylation in vivo induces autoimmune disease. Proc. Natl. Acad. Sci. USA *98*, 1142–1147.

Culyba, E.K., Price, J.L., Hanson, S.R., Dhar, A., Wong, C.H., Gruebele, M., Powers, E.T., and Kelly, J.W. (2011). Protein native-state stabilization by placing aromatic side chains in N-glycosylated reverse turns. Science *331*, 571–575.

Freeze, H.H. (2006). Genetic defects in the human glycome. Nat. Rev. Genet. *7*, 537–551.

Gavel, Y., and von Heijne, G. (1990). Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites: implications for protein engineering. Protein Eng. *3*, 433–442.

Glozman, R., Okiyoneda, T., Mulvihill, C.M., Rini, J.M., Barriere, H., and Lukacs, G.L. (2009). N-glycans are direct determinants of CFTR folding and stability in secretory and endocytic membrane traffic. J. Cell Biol. *184*, 847–862.

Hammarstrom, P., Sekijima, Y., White, J.T., Wiseman, R.L., Lim, A., Costello, C.E., Altland, K., Garzuly, F., Budka, H., and Kelly, J.W. (2003). D18G transthyretin is monomeric, aggregation prone, and not detectable in plasma and cerebrospinal fluid: a prescription for central nervous system amyloidosis? Biochemistry *42*, 6656–6663.

Hammond, C., Braakman, I., and Helenius, A. (1994). Role of N-linked oligosaccharide recognition, glucose trimming, and calnexin in glycoprotein folding and quality control. Proc. Natl. Acad. Sci. USA *91*, 913–917.

Hanson, S.R., Culyba, E.K., Hsu, T.L., Wong, C.H., Kelly, J.W., and Powers, E.T. (2009). The core trisaccharide of an N-linked glycoprotein intrinsically accelerates folding and enhances stability. Proc. Natl. Acad. Sci. USA *106*, 3131–3136.

Hebert, D.N., Lamriben, L., Powers, E.T., and Kelly, J.W. (2014). The intrinsic and extrinsic effects of N-linked glycans on glycoproteostasis. Nat. Chem. Biol. *10*, 902–910.

Helenius, A., and Aebi, M. (2001). Intracellular functions of N-linked glycans. Science *291*, 2364–2369.

Hsieh, P., Rosner, M.R., and Robbins, P.W. (1983). Selective cleavage by endo-beta-N-acetylglucosaminidase H at individual glycosylation sites of Sindbis virion envelope glycoproteins. J. Biol. Chem. *258*, 2555–2561.

Hubbard, S.C. (1988). Regulation of glycosylation. The influence of protein structure on N-linked oligosaccharide processing. J. Biol. Chem. *263*, 19303–19317.

Imperiali, B., and Rickert, K.W. (1995). Conformational implications of asparagine-linked glycosylation. Proc. Natl. Acad. Sci. USA *92*, 97–101.

Jaeken, J. (2010). Congenital disorders of glycosylation. Ann. N. Y. Acad. Sci. *1214*, 190–198.

Jang-Lee, J., North, S.J., Sutton-Smith, M., Goldberg, D., Panico, M., Morris, H., Haslam, S., and Dell, A. (2006). Glycomic profiling of cells and tissues by mass spectrometry: fingerprinting and sequencing methodologies. Methods Enzymol. *415*, 59–86.

Jitsuhara, Y., Toyoda, T., Itai, T., and Yamaguchi, H. (2002). Chaperone-like functions of high-mannose type and complex-type N-glycans and their molecular basis. J. Biochem. *132*, 803–811.

Joao, H.C., and Dwek, R.A. (1993). Effects of glycosylation on protein structure and dynamics in ribonuclease B and some of its individual glycoforms. Eur. J. Biochem. *218*, 239–244.

 ©2015 Elsevier Ltd All rights reserved

Kasturi, L., Eshleman, J.R., Wunner, W.H., and Shakin-Eshleman, S.H. (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.

Kasturi, L., Chen, H., and Shakin-Eshleman, S.H. (1997). Regulation of N-linked core glycosylation: use of a site-directed mutagenesis approach to identify Asn-Xaa-Ser/Thr sequons that are poor oligosaccharide acceptors. Biochem. J. *323*, 415–419.

Kornfeld, R., and Kornfeld, S. (1985). Assembly of asparagine-linked oligosaccharides. Annu. Rev. Biochem. *54*, 631–664.

Kowarik, M., Young, N.M., Numao, S., Schulz, B.L., Hug, I., Callewaert, N., Mills, D.C., Watson, D.C., Hernandez, M., Kelly, J.F., et al. (2006). Definition of the bacterial N-glycosylation site consensus sequence. EMBO J. *25*, 1957–1966.

Lowe, J.B., and Marth, J.D. (2003). A genetic approach to mammalian glycan function. Annu. Rev. Biochem. *72*, 643–691.

Lu, X., Mehta, A., Dadmarz, M., Dwek, R., Blumberg, B.S., and Block, T.M. (1997). Aberrant trafficking of hepatitis B virus glycoproteins in cells in which N-glycan processing is inhibited. Proc. Natl. Acad. Sci. USA *94*, 2380–2385.

Martina, J.A., Daniotti, J.L., and Maccioni, H.J. (1998). Influence of N-glycosylation and N-glycan trimming on the activity and intracellular traffic of GD3 synthase. J. Biol. Chem. *273*, 3725–3731.

Morrison, K.L., and Weiss, G.A. (2001). Combinatorial alanine-scanning. Curr. Opin. Chem. Biol. *5*, 302–307.

Nairn, A.V., York, W.S., Harris, K., Hall, E.M., Pierce, J.M., and Moremen, K.W. (2008). Regulation of glycan structures in animal tissues: transcript profiling of glycan-related genes. J. Biol. Chem. *283*, 17298–17313.

Oliver, J.D., van der Wal, F.J., Bulleid, N.J., and High, S. (1997). Interaction of the thiol-dependent reductase ERp57 with nascent glycoproteins. Science *275*, 86–88.

Ou, W.J., Cameron, P.H., Thomas, D.Y., and Bergeron, J.J. (1993). Association of folding intermediates of glycoproteins with calnexin during protein maturation. Nature *364*, 771–776.

Parodi, A.J. (2000). Role of N-oligosaccharide endoplasmic reticulum processing reactions in glycoprotein folding and degradation. Biochem. J. *348*, 1–13.

Petrescu, A.J., Milac, A.L., Petrescu, S.M., Dwek, R.A., and Wormald, M.R. (2004). Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. Glycobiology *14*, 103–114.

Price, J.L., Powers, D.L., Powers, E.T., and Kelly, J.W. (2011). Glycosylation of the enhanced aromatic sequon is similarly stabilizing in three distinct reverse turn contexts. Proc. Natl. Acad. Sci. USA *108*, 14127–14132.

Price, J.L., Culyba, E.K., Chen, W., Murray, A.N., Hanson, S.R., Wong, C.H., Powers, E.T., and Kelly, J.W. (2012). N-glycosylation of enhanced aromatic sequons to increase glycoprotein stability. Biopolymers *98*, 195–211.

Recny, M.A., Luther, M.A., Knoppers, M.H., Neidhardt, E.A., Khandekar, S.S., Concino, M.F., Schimke, P.A., Francis, M.A., Moebius, U., Reinhold, B.B., et al. (1992). N-glycosylation is required for human CD2 immunoadhesion functions. J. Biol. Chem. *267*, 22428–22434.

Rudd, P.M., Wormald, M.R., Harvey, D.J., Devasahayam, M., McAlister, M.S., Brown, M.H., Davis, S.J., Barclay, A.N., and Dwek, R.A. (1999). Oligosaccharide analysis and molecular modeling of soluble forms of glycoproteins belonging to the Ly-6, scavenger receptor, and immunoglobulin superfamilies expressed in Chinese hamster ovary cells. Glycobiology *9*, 443–458.

Ruiz-Canada, C., Kelleher, D.J., and Gilmore, R. (2009). Cotranslational and posttranslational N-glycosylation of polypeptides by distinct mammalian OST isoforms. Cell *136*, 272–283.

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

Schmaltz, R.M., Hanson, S.R., and Wong, C.H. (2011). Enzymes in the synthesis of glycoconjugates. Chem. Rev. *111*, 4259–4307.

Sekijima, Y., Wiseman, R.L., Matteson, J., Hammarstrom, P., Miller, S.R., Sawkar, A.R., Balch, W.E., and Kelly, J.W. (2005). The biological and chemical basis for tissue-selective amyloid disease. Cell *121*, 73–85.

Shrimal, S., Ng, B.G., Losfeld, M.E., Gilmore, R., and Freeze, H.H. (2013a). Mutations in STT3A and STT3B cause two congenital disorders of glycosylation. Hum. Mol. Genet. *22*, 4638–4645.

Shrimal, S., Trueman, S.F., and Gilmore, R. (2013b). Extreme C-terminal sites are posttranslocationally glycosylated by the STT3B isoform of the OST. J. Cell Biol. *201*, 81–95.

Sorensen, A.L., Clausen, H., and Wandall, H.H. (2012). Carbohydrate clearance receptors in transfusion medicine. Biochim. Biophys. Acta *1820*, 1797–1808.

Surleac, M.D., Spiridon, L.N., Tacutu, R., Milac, A.L., Petrescu, S.M., and Petrescu, A.J. (2012). The structural assessment of glycosylation sites database - SAGS - an overall view on N-glycosylation. In Glycosylation, S.M. Petrescu, ed. (InTech), pp. 3–20.

Thaysen-Andersen, M., and Packer, N.H. (2012). Site-specific glycoproteomics confirms that protein structure dictates formation of N-glycan type, core fucosylation and branching. Glycobiology *22*, 1440–1452.

Trimble, R.B., Maley, F., and Chu, F.K. (1983). Glycoprotein biosynthesis in yeast. protein conformation affects processing of high mannose oligosaccharides on carboxypeptidase Y and invertase. J. Biol. Chem. *258*, 2562–2567.

Vembar, S.S., and Brodsky, J.L. (2008). One step at a time: endoplasmic reticulum-associated degradation. Nat. Rev. Mol. Cell Biol. *9*, 944–957.

Wacker, M., Linton, D., Hitchen, P.G., Nita-Lazar, M., Haslam, S.M., North, S.J., Panico, M., Morris, H.R., Dell, A., Wren, B.W., et al. (2002). N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. Science *298*, 1790–1793.

Wada, Y., Azadi, P., Costello, C.E., Dell, A., Dwek, R.A., Geyer, H., Geyer, R., Kakehi, K., Karlsson, N.G., Kato, K., et al. (2007). Comparison of the methods for profiling glycoprotein glycans—HUPO Human Disease Glycomics/Proteome Initiative multi-institutional study. Glycobiology *17*, 411–422.

Wang, C., Eufemi, M., Turano, C., and Giartosio, A. (1996). Influence of the carbohydrate moiety on the stability of glycoproteins. Biochemistry *35*, 7299–7307.

Ware, F.E., Vassilakos, A., Peterson, P.A., Jackson, M.R., Lehrman, M.A., and Williams, D.B. (1995). The molecular chaperone calnexin binds Glc1Man9GlcNAc2 oligosaccharide as an initial step in recognizing unfolded glycoproteins. J. Biol. Chem. *270*, 4697–4704.

Wormald, M.R., and Dwek, R.A. (1999). Glycoproteins: glycan presentation and protein-fold stability. Structure *7*, R155–R160.

Wyss, D.F., Choi, J.S., Li, J., Knoppers, M.H., Willis, K.J., Arulanandam, A.R., Smolyar, A., Reinherz, E.L., and Wagner, G. (1995). Conformation and function of the N-linked glycan in the adhesion domain of human CD2. Science *269*, 1273–1278.