Single-particle electron microscopy structure of UDP-glucose:glycoprotein glucosyltransferase suggests a selectivity mechanism for misfolded proteins

Received for publication, April 3, 2017, and in revised form, May 7, 2017 Published, Papers in Press, May 10, 2017, DOI 10.1074/jbc.M117.789495

Daniel Calles-Garcia[‡1], Meng Yang[‡1], Naoto Soya[§], Roberto Melero[¶], Marie Ménade[‡], Yukishige Ito[‖], Javier Vargas[¶], Gergely L. Lukacs[§], Justin M. Kollman[‡‡], Guennadi Kozlov[‡], and Kalle Gehring[‡2]**

*From the [‡]Department of Biochemistry, McGill University, Montreal, Quebec H3G0B1, Canada, [§]Department of Physiology, McGill University, Montreal, Quebec H3G1Y6, Canada, [¶]Biocomputing Unit, Centro Nacional de Biotectnologíay, 28049 Madrid, Spain, [‖]Synthetic Cellular Chemistry Laboratory, RIKEN, Wako, Saitama 351-0198, Japan, **Bioengineering Lab, Escuela Politécnica Superior, Universidad San Pablo CEU, 28668 Madrid, Spain, and [‡‡]Department of Biochemistry, University of Washington, Seattle, Washington 98195-7350*

Edited by Norma Allewell

The enzyme UDP-glucose:glycoprotein glucosyltransferase (UGGT) mediates quality control of glycoproteins in the endoplasmic reticulum by attaching glucose to *N*-linked glycan of misfolded proteins. As a sensor, UGGT ensures that misfolded proteins are recognized by the lectin chaperones and do not leave the secretory pathway. The structure of UGGT and the mechanism of its selectivity for misfolded proteins have been unknown for 25 years. Here, we used negative-stain electron microscopy and small-angle X-ray scattering to determine the structure of UGGT from *Drosophila melanogaster* at 18-Å resolution. Three-dimensional reconstructions revealed a cage-like structure with a large central cavity. Particle classification revealed flexibility that precluded determination of a high-resolution structure. Introduction of biotinylation sites into a fungal UGGT expressed in *Escherichia coli* allowed identification of the catalytic and first thioredoxin-like domains. We also used hydrogen-deuterium exchange mass spectrometry to map the binding site of an accessory protein, Sep15, to the first thioredoxin-like domain. The UGGT structural features identified suggest that the central cavity contains the catalytic site and is lined with hydrophobic surfaces. This enhances the binding of misfolded substrates with exposed hydrophobic residues and excludes folded proteins with hydrophilic surfaces. In conclusion, we have determined the UGGT structure, which enabled us to develop a plausible functional model of the mechanism for UGGT's selectivity for misfolded glycoproteins.

The folding of glycoproteins that transit through the endoplasmic reticulum (ER)[3] is reflected in the structure of their *N*-linked glycan. Newly synthesized proteins are modified with the intact glycan $Glc_3Man_9GlcNAc_2$, which is then trimmed by glucosidases I and II to generate a monoglucosylated form. This trimmed form is recognized by the ER-specific lectin chaperones calnexin and calreticulin, which recruit additional chaperones to fold the nascent protein (1–5). Upon release, the terminal glucose is cleaved by glucosidase II, producing a shortened glycan, $Man_9GlcNAc_2$, which can no longer bind the lectin chaperones. Incompletely folded or misfolded proteins after the first cycle of folding are recognized by UDP-glucose:glycoprotein glucosyltransferase (UGGT), which adds back a glucose residue to regenerate the monoglucosylated form for additional rounds of lectin chaperone-assisted refolding (6, 7). By glucosylating misfolded proteins, UGGT plays an important role in preventing misfolded proteins from exiting the ER.

UGGT is conserved and essential (8, 9). It is a large monomeric protein of more than 1500 residues and found in almost all eukaryotes. The activity of UGGT has been probed using native and misfolded glycoproteins (7, 10–13), glycopeptides (14–16), and small synthetic substrates (17–22). These studies have shown a strong selectivity for glucosylation of misfolded over folded substrates. Only glycans in misfolded proteins are modified. UGGT activity requires that misfolded protein and the acceptor glycan $Man_9GlcNAc_2$ are on the same molecule (23, 24). UGGT is known to tightly bind another ER protein, Sep15, a 15-kDa selenoprotein that likely contributes to the reduction of non-native disulfide bonds (25–27).

In addition to its role as a sensor of protein folding, UGGT contributes to the maturation of the major histocompatibility complex of class 1 molecules (MHC1), ensuring that a high-affinity peptide has been loaded onto the MHC1 heavy chains (28, 29). The MHC1 is composed of MHC-encoded heavy chains, which are glycosylated, and a β_2-microglobulin domain. The assembly of the complete MHC1 involves many components of the ER folding pathways. UGGT scans the peptide-MHC1 interaction through glucosylating the heavy chain and directs the MHC1 for reassociation with the calreticulin-

This work was supported by Natural Sciences and Engineering Research Council of Canada Grant RPGIN 2014-04686 (to K. G.). The authors declare that they have no conflicts of interest with the contents of this article.

This article contains supplemental Figs. S1–S14 and Movies S1 and S2.

[1] Both authors contributed equally to this work.

[2] To whom correspondence should be addressed. Tel.: 514-398-7287; E-mail: kalle.gehring@mcgill.ca.

[3] The abbreviations used are: ER, endoplasmic reticulum; UGGT, UDP-glucose:glycoprotein glucosyltransferase; DmUGGT, *D. melanogaster* UGGT; PcUGGT, *P. chrysogenum* UGGT; SAXS, small-angle X-ray scattering; HDX, hydrogen-deuterium exchange; mSA, monovalent streptavidin; m, mutant; Ni-NTA, nickel-nitrilotriacetic acid; CCD, charge-coupled device; FSC, Fourier shell correlation; SEC, size-exclusion chromatography; UHPLC, ultra-HPLC.

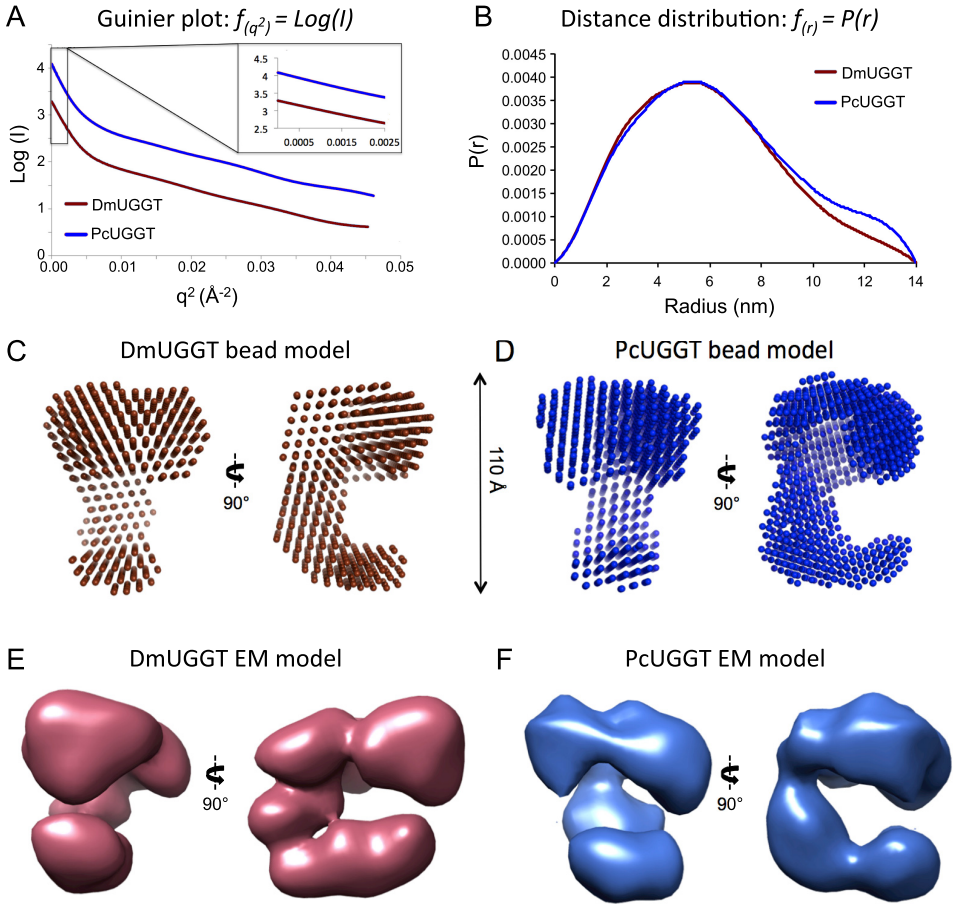

Figure 1. SAXS analysis of UGGT. *A*, Guinier analysis of data showed that the proteins were monomeric and not aggregated. *B*, distance distribution plots from the DmUGGT and PcUGGT scattering data calculated by GNOM (34) with a maximum diameter of 140 Å. *C* and *D*, *ab initio* structures calculated from data processed with a D_{max} of 100 Å. *E* and *F*, for comparison, the EM models of DmUGGT and PcUGGT calculated from EMAN are shown.

ERp57-tapasin complex for a new peptide to be loaded onto the heavy chain-binding groove. UGGT ensures that only high-affinity peptides are loaded onto the MHC1.

A mechanistic understanding of UGGT activity has been limited as the protein has proven to be a difficult crystallization target. Sequence analysis and biochemical data have suggested that it has five domains. The N-terminal portion consists of three thioredoxin-like domains followed by a β-strand-rich domain and the catalytic glucosyltransferase domain. The catalytic domain is the most strongly conserved across species (9, 30). The non-catalytic domains are less conserved with 20–40% sequence identity among species. The only crystal structure is of the third thioredoxin-like domain from *Chaetomium thermophilum* (31).

In this report, we characterize the structure of UGGT from *Drosophila melanogaster* and *Penicillium chrysogenum*. Solution analysis by small-angle X-ray and light scattering confirmed that *D. melanogaster* UGGT (DmUGGT) is monomeric and forms a complex with Sep15 in a 1:1 ratio. We identified the Sep15-binding site on UGGT using hydrogen-deuterium exchange with mass spectrometry (HDX-MS) and mutagenesis. Small-angle X-ray scattering (SAXS) and single-particle electron microscopy (EM) revealed that UGGT adopts an open cage shape with a large central cavity. We identified the domains in an EM model of *P. chrysogenum* UGGT (PcUGGT) using biotinylation sites decorated with monovalent streptavidin. The three-dimensional model of UGGT allowed us to propose a mechanism for substrate recognition and selection.

Results

SAXS analysis

We chose to study UGGT from two sources. DmUGGT had been previously characterized biochemically and is easily purified from Sf9 insect cells due to incorporation of an N-terminal melittin signal peptide that secretes the protein into the culture medium (2, 32, 33). We also studied the fungal PcUGGT, which could be expressed in *Escherichia coli* and purified in high yield (supplemental Fig. S1). The two proteins are 44 and 34% identical to human UGGT with the strongest similarity in the C-terminal catalytic domain.

To characterize the proteins, we turned to SAXS and EM, two structural techniques that have been used successfully to study proteins that are difficult to crystallize. SAXS analysis of DmUGGT and PcUGGT confirmed that the proteins were monomeric with no aggregation detected in Guinier plots (Fig. 1*A*). Kratky plots of the scattering data showed an elongated bell curve that returned to zero, which is characteristic of a folded protein (supplemental Fig. S2). The radius of gyration for Dm- and PcUGGT, calculated from the slope of the Guinier

Table 1
EM data collection and analysis

Sample	Magnification	Sampling	Defocus	Particles analyzed	Resolution		Software
					$FSC_{0.5}$	$FSC_{0.143}$	
		Å/pixel	*μm*		*Å*		
DmUGGT	62,000	1.8	−2 to −4	26,624	17.8	12.5	Scipion/RELION2.0
PcUGGT	62,000	1.8	−2 to −4	32,712	33.3	18.2	EMAN2.0
PcUGGT-biotin-mSA-m3	62,000	1.8	−1.5 to −3.5	12,663	20.5	15.2	EMAN2.0/Scipion
PcUGGT-biotin-mSA-m15	62,000	1.8	−1.5 to −3.5	10,743	22.9	16.7	EMAN2.0

plots at low angle, was estimated to 4.9 nm and independent of protein concentration. The distance distribution curves for both species showed similar bell-shaped curves consistent with D_{max} values in the range of 100–140 Å (Fig. 1*B*).

We used the SAXS data to generate *ab initio* bead models for DmUGGT and PcUGGT using the ATSAS suite of programs (34). The two models were very similar and displayed a C-shaped structure with two distinct regions or lobes (Fig. 1, *C* and *D*). The larger lobe was 80 × 70 × 50 Å with a depression at its center and was connected to an elongated lobe of 70 × 40 × 40 Å. The distance separating the lobes was strongly dependent on the D_{max} used for the structure calculations. The structures calculated with larger D_{max} values showed a T-shaped conformation, whereas in the smaller D_{max} structures the elongated lobe was curved with a distance of roughly 35 Å between the ends of the lobes (supplemental Fig. S3).

EM analysis

We next carried out negative-stain EM studies of DmUGGT and PcUGGT (Table 1). Single-particle analysis using uranyl formate stain showed that the protein particles had no preferred orientation on the grid, and the size distribution of the particles was very homogeneous, key characteristics to facilitate three-dimensional reconstructions. Initial analysis was done in EMAN2.0 (35, 36) to create two-dimensional reference-free class averages. The resulting models of DmUGGT and PcUGGT after initial model constructions and three-dimensional refinements showed high consistencies and similarities. Back-projections of the three-dimensional maps agreed with the two-dimensional class averages observed experimentally. The refined models, determined to a resolution of ~25 Å using EMAN2.0, exhibited a distinctive C-shape when viewed from the side and a ring shape from the top (Fig. 1 and supplemental Figs. S4 and S5).

To confirm our EM model, we processed the data using different software. Briefly, the boxed particles of DmUGGT were re-extracted and further selected through two-dimensional alignments in RELION2.0 (37). The new particle subset was imported into Scipion (38) for further two-dimensional classification using RELION and CL2D (39). Initial volumes were estimated using a different method (40) followed by RELION three-dimensional classifications and RELION autorefinement (Fig. 2, *A* and *B*). No significant differences were observed in the classes obtained using the different processing software, but the DmUGGT structure determined using Scipion provided slightly improved resolution (Fig. 2*C*). The higher resolution structure confirmed the claw-like architecture composed of a ring-shaped lobe connected to an arm-like lobe (Fig. 2, *B* and *D*).

As an additional check, we calculated the theoretical X-ray scattering curve from the Scipion DmUGGT EM model and compared it with the experimental data (supplemental Fig. S6). A good fit was observed at small angles ($q < 0.06$ Å^{-1}), but the curves diverged at larger angles, which could result from distortion during dehydration of the EM samples or conformational heterogeneity. To address the latter, we re-examined the Scipion three-dimensional classifications and detected the presence of multiple particle classes (Fig. 2*D*). The classes were very similar but showed distinct and reproducible position and density differences. We detected multiple classes irrespective of the classification parameters used, which suggests that flexibility is an intrinsic property of the protein. To visualize the differences in the structures, we generated morphings between the classes (supplemental Movies 1 and 2). The DmUGGT classes were evenly populated with small incremental changes, suggesting continuous movement or flexibility in the structure. Despite these differences, the organization of UGGT as two lobes with a central cavity was consistent across all structural models (Fig. 2*D*).

Domain identification

To identify the features in the EM and SAXS maps, we used the DOLORS strategy of labeling with monovalent streptavidin (mSA) using inserted biotinylation sites (41). We chose to apply the strategy to PcUGGT because its facile expression allowed screening of a large number of insertions. Building on the bioinformatics analysis of UGGT from *C. thermophilum* (31), we examined the primary sequence of PcUGGT and selected 16 putative loop regions for insertion of biotinylation sites (Fig. 3*A*). After expression, purification, and mSA labeling, we obtained five PcUGGT-biotin-mSA samples for EM analysis (supplemental Fig. S7).

Two of the streptavidin-labeled PcUGGT samples gave clearly defined additional density that allowed us to assign parts of the EM map to specific domains. PcUGGT-biotin-mSA-m3 (mutant 3) contained a biotinylation site inserted after lysine 261 in the first thioredoxin domain, and PcUGGT-biotin-mSA-m15 (mutant 15) contained a biotinylation site inserted after aspartic acid 1377 in the catalytic domain (supplemental Fig. S7). The reference-free classifications of the boxed particles of the streptavidin-labeled mutants were very similar to the classes observed for PcUGGT with additional density consistent in size with a monovalent streptavidin label (Fig. 3, *B* and *C*). The structures were determined to resolutions of 20 Å (Table 1 and supplemental Fig. S8).

The EM map of PcUGGT-biotin-mSA-m3 showed additional density on the ring lobe and identifies this as the N-terminal portion of UGGT (Fig. 3*D*). We posit that the ring is

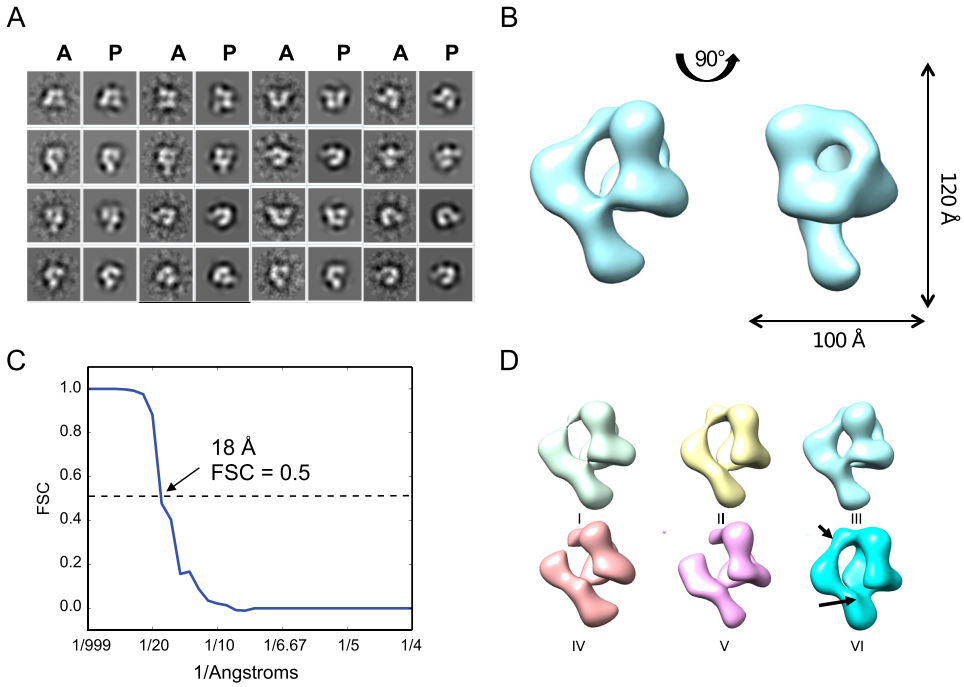

Figure 2. *A*, two-dimensional class averages from Xmipp CL2D (*A*) and the back-projections (*P*) calculated from the final three-dimensional map generated in Scipion. *B*, EM model of DmUGGT from single-particle analysis of negative-stain images with Scipion. The density was thresholded at 0.05 in UCSF Chimera. *C*, FSC curve of the map. *D*, the three-dimensional classes simultaneously exist in equal abundance from classes I–VI. The major differences are indicated by *arrows* in class VI.

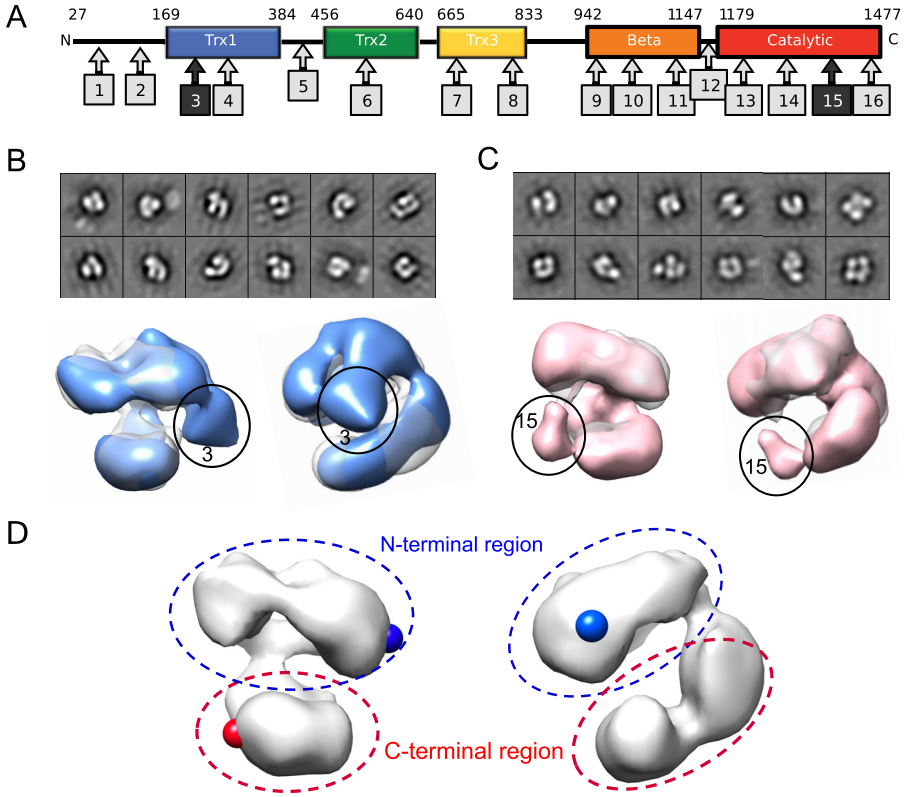

Figure 3. *A*, the domain structure of PcUGGT and location of the 16 biotinylation sites introduced. EM structures were determined for mutants 3 and 15. *B*, two-dimensional class averages and negative-stain EM map of monovalent streptavidin-labeled PcUGGT mutant 3 (*blue*; threshold level, 0.05) overlaid with the wild-type map. The extra density from streptavidin is *circled*. *C*, two-dimensional class averages and negative-stain EM map of monovalent streptavidin-labeled PcUGGT mutant 15 (*pink*; threshold level, 1.67) overlaid on the wild-type map. *D*, the streptavidin attachment sites for mutants 3 (*blue*) and 15 (*red*) are shown on the wild-type PcUGGT EM map (threshold level, 1.38).

formed by the three thioredoxin-like domains and that the arm-like lobe is formed by the β-strand-rich and catalytic domains. In agreement with this, the EM map of PcUGGT-biotin-mSA-m15 showed distinct additional density at the end of the arm-like lobe. Localization of the catalytic domain at the end of the arm allows us to determine that the β-strand-rich domain is in

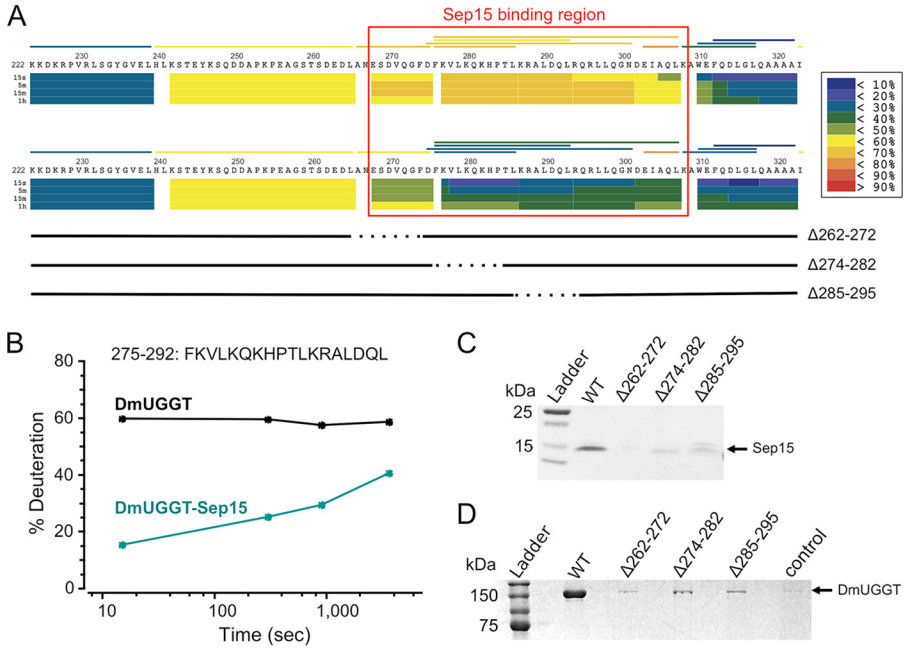

Figure 4. *A*, HDX-MS results for DmUGGT and the DmUGGT-Sep15 complex. *Blue* (0%) to *red* (100%) indicates the extent of deuterium exchange. For clarity, only the region around the Sep15-binding site is shown. *B*, deuteration of a peptide in the Sep15-binding site measured at time points between 15 s and 1 h. In the absence of Sep15, the peptide was completely exchanged at the first time point. *C*, SDS-PAGE analysis of the binding between DmUGGT mutants and Sep15 measured by co-elution in size-exclusion chromatography. DmUGGT and Sep15 are present in the complex in equal molar amounts. *D*, pulldown of DmUGGT by Sep15 immobilized on CNBr-activated Sepharose beads. Beads without Sep15 served as the control.

the density connecting the catalytic domain and N-terminal lobe.

Identification of Sep15-binding site

In cells, UGGT is tightly associated with Sep15, a small 15-kDa thioredoxin-like protein. Isothermal titration calorimetry experiments measured the affinity to be 20 nM (27). To characterize the site of its interaction on UGGT, we expressed and purified Sep15 from *E. coli* cells. The complex of DmUGGT and Sep15 was prepared by mixing an excess of Sep15 with purified DmUGGT followed by gel filtration chromatography to separate the complex from excess Sep15. Formation of the complex could be observed by a shift in the elution time, an increase in the molecular weight measured by multiangle light scattering, and SDS-PAGE analysis (supplemental Fig. S9).

We utilized HDX-MS to identify the Sep15-binding site on DmUGGT. The technique is based on measuring differences in the exchange of amide hydrogens in complex and free protein. The peptide coverage in MS analysis of DmUGGT was excellent, allowing measurement of the solvent accessibility of 95% of the protein (supplemental Fig. S10). Comparison of the exchange rates for DmUGGT and DmUGGT-Sep15 complex showed a marked difference in a roughly 40-residue region predicted to be in the first thioredoxin-like domain (Fig. 4*A*). Residues in the region displayed a significantly reduced rate of hydrogen-deuterium exchange in the presence of Sep15 (Fig. 4*B* and supplemental Fig. S11). Outside of this region, there were no significant differences (supplemental Figs. S12 and S13).

To confirm the identification of the Sep15-binding site, we prepared three DmUGGT mutants with deletions within the Sep15-binding region: DmUGGT-Δ262–272, DmUGGT-Δ274–282, and DmUGGT-Δ285–295. A fourth construct, DmUGGT-Δ298–305, showed significant protein degradation, suggesting that the deletion destabilized the protein structure. Sep15 binding to the mutants was tested by co-elution on a gel filtration column. Analysis by SDS-PAGE showed that the DmUGGT-Δ262–272 mutant had completely lost the ability to bind Sep15, and the mutants DmUGGT-Δ274–282 and DmUGGT-Δ285–295 were strongly impaired (Fig. 4*C*). Identical results were obtained by using Sep15 coupled to cyanogen bromide (CNBr)-activated Sepharose beads to pull down wild-type and mutant DmUGGT proteins (Fig. 4*D* and supplemental Fig. S14).

Discussion

UGGT has been the target of numerous structural efforts over the last 25 years. Here, we used EM and SAXS to characterize its overall shape. The main features of the structure are very robust. We observed highly similar conformation for UGGT from two evolutionarily distant organisms with only 33% amino acid sequence identity. It is also notable that the structures calculated from the SAXS and EM data are very similar. The back-projections of the SAXS models in Fig. 1 agree well with the EM two-dimensional class averages in Fig. 2, and the overall three-dimensional shapes are similar. *Ab initio* SAXS modeling is not well suited for the analysis of hollow structures such as UGGT, but it nonetheless provided a strong independent confirmation of the EM maps.

The identification of the Sep15-binding site is similarly robust. The HDX-MS data are remarkably clear and identify a highly charged segment in the first thioredoxin-like domain as the binding site. The region is loosely structured in the absence of Sep15 as determined by its complete deuteration at 15 s.

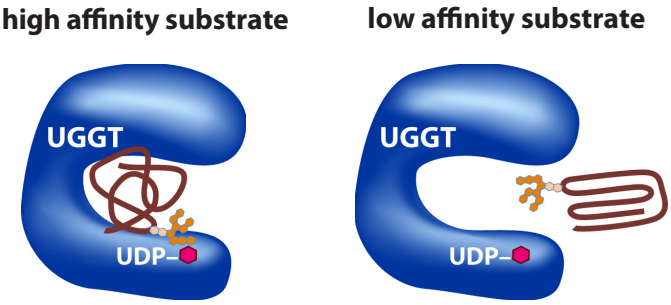

Figure 5. Model for the selectivity of UGGT for unfolded proteins. Hydrophobic surfaces in the central cavity of UGGT bind misfolded proteins to increase their concentration at the catalytic site. Folded proteins are excluded due to their hydrophilicity and size.

Homology modeling suggests the segment could adopt an α-helix upon binding of the Sep15. Sep15 is composed of a cysteine-rich domain that binds UGGT and a selenocysteine-containing thioredoxin-like domain (25–27). The protein is highly conserved across eukaryotes and thought to act as a protein-disulfide reductase due to the oxidation potential of its selenocysteine (26).

Although limited to 20-Å resolution, the UGGT structure is surprisingly informative and suggests a mechanism for substrate recognition. A large number of studies have characterized the specificity of UGGT for misfolded proteins (7, 10–22). Full activity requires both exposed hydrophobic residues and the $Man_9GlcNAc_2$ glycan, which is the site of monoglucosylation and release. The two elements have to be in the same molecule, and in fact addition of deglycosylated (endoglycosidase H-treated) misfolded proteins inhibits UGGT activity (42). There is a bias toward glycans adjacent to hydrophobic residues and against bulky folded proteins. Studies with synthetic substrates composed of the acceptor glycan linked to methotrexate have shown that the addition of dihydrofolate reductase reduces activity (17). Allosteric coupling between a domain that acts a sensor of misfolded protein and the glucosyltransferase domain has been proposed, although it fails to explain the requirement for both recognition elements to be on the same molecule.

The UGGT structure suggests a simpler mechanism for the selection of misfolded proteins that is consistent with the literature and involves minimal complexity. We hypothesize that the catalytic site is located inside the UGGT cavity to sequester the glucosyltransferase activity away from glycans on folded, hydrophilic substrates (Fig. 5). We expect that the cavity will be lined with hydrophobic surfaces to favor the recruitment of misfolded proteins that have exposed hydrophobic residues. In a manner similar to the chaperonins GroEL/GroES, UGGT recruits substrates into a hydrophobic cavity. Misfolded proteins are kinetically favored through their higher rate of occupancy in the cavity. The size of the pocket favors the binding of intact, partially folded proteins over small hydrophobic glycopeptides (43), and flexibility allows UGGT to act on different misfolded proteins. This simple mechanism explains why the acceptor glycan and exposed hydrophobic segment need to be in the same molecule and why misfolded, deglycosylated proteins inhibit UGGT activity.

Testing of this model will require a higher resolution structure of UGGT to identify the nature of internal surfaces of the cavity and the orientation of catalytic domain. It is interesting that the recently determined structure of the third thioredoxin-like domain from *C. thermophilum* UGGT revealed the existence of the extensive detergent-bound hydrophobic patch (31). Future studies are necessary to determine whether the other thioredoxin-like domains in the N-terminal lobe similarly contain hydrophobic surfaces.

Experimental procedures

DmUGGT expression and purification

DmUGGT cloned into the pFastBac1 plasmid with a N-terminal melittin secretion sequence and a C-terminal hexahistidine tag was obtained from Daniel Tessier (2). The reconstructed plasmid was transformed into DH10Bac *E. coli* cells, and the baculovirus DNA was produced according to the Sf9 baculovirus expression system manual using serum-free media. Sf9 cells were amplified to 2×10^6/ml and infected by P3 virus with a ratio of 0.2% (v/v). The Sf9 cells continued to grow for 72 h at 27 °C at 85 rpm. The supernatant containing the secreted DmUGGT was separated with the cells by centrifugation at 1000 rpm for 15 min. The supernatant was loaded into the pre-equilibrated Ni-NTA gravity column. After all the supernatant flowed through, the column was washed using Buffer A (30 mM Tris-HCl, pH 7.5, 300 mM NaCl, 20 mM imidazole). The protein was eluted from the column using Buffer B (30 mM Tris-HCl, pH 7.5, 300 mM NaCl, 300 mM imidazole). DmUGGT was concentrated to less than 5 ml using a 50-kDa-molecular-mass-cutoff Amicon filter by centrifugation at 4000 rpm at 4 °C before being injected into a Superdex 200 prep grade HiLoad 16/600 column, which was pre-equilibrated with Buffer C (30 mM Tris-HCl, pH 7.5, 100 mM NaCl). The fractions were analyzed by SDS-PAGE.

Expression and purification of PcUGGT

P. chrysogenum UGGT was codon-optimized and cloned into pCold I vector for bacterial expression using *E. coli* Rosetta Gami 2. Cells were grown until $A_{600\ nm}$ reaches 0.6 and then induced with 0.5 mM isopropyl 1-thio-β-D-galactopyranoside at 37 °C for 4 h. Cells were harvested by centrifugation, suspended in lysis buffer (Buffer A), and lysed by sonication. The cell lysate was centrifuged at 18,000 rpm at 4 °C for 60 min, and the supernatant was subjected to Ni-NTA purification. The eluted sample was further purified by anion-exchange chromatography and size-exclusion chromatography. The fractions were analyzed by SDS-PAGE.

SAXS data collection and analysis

DmUGGT at 0.5–20 mg/ml was used for data collection on an in-house SAXSess system (Anton Paar) with exposures of 6 h per sample on a CCD chip for both protein and buffer samples. The SAXS data were processed and analyzed using the ATSAS software package (34). PRIMUS was used to subtract buffer scattering curves and normalize the data by protein concentration. GNOM was used to calculate $P(r)$ curves with different D_{max} values. For each D_{max} value, 20 pseudoatom (bead) mod-

els were generated by DAMMIF and then aligned and compared in DAMSEL to determine the most probable model. DAMSUP was used to align the models with the most probable model, and the aligned models were averaged and filtered in DAMAVER to generate the final SAXS structure. For comparison of the SAXS and EM data, EM2DAM was used to calculate a theoretical X-ray scattering curve from an input EM model.

EM sample preparation and image acquisition

DmUGGT (12 ng/μl) and PcUGGT (7.5 ng/μl) proteins were applied onto negatively glow-discharged carbon-coated grids (400-mesh copper grids) for 60 s, and excess protein was removed by blotting using filter papers. Freshly prepared 0.75% uranyl formate, pH 5, was applied on the grids after the protein was removed for 60 s and then blotted. The grids were dried for at least 30 min before electron microscopy. Images were collected using an FEI Tecnai G2 TF20 transmission electron microscope at 200 kV equipped with a Gatan Ultrascan 4000 CCD camera (model 895) at 62,000× magnification and a 1.8-Å/pixel sampling rate. Micrographs were recorded at varying defocus values (Table 1).

EM data processing and image analysis

For all the data sets, protein particles were picked using e2boxer from EMAN2.0 with a 140 × 140-pixel box size. The particles were either extracted in EMAN2.0 (35, 36) or extracted in RELION2.0 (37). Contrast transfer function corrections were done using CTFFind3. For EMAN2.0 processing, particles were classified using reference-free algorithm *K*-means, and only particles from well defined class averages were kept for further processing. The high-quality particles were used to generate initial models by the common line algorithm within EMAN2.0 and by the SIMPLE PRIME (44, 45) method. We refined these models within EMAN2.0 using the gold-standard Fourier shell correlation method. UGGT models were compared using Chimera at thresholds levels giving equivalent voxel volumes. Back-projections of the models were calculated using EMAN2.0 e2project3d.py for comparison with the class averages (supplemental Figs. S4 and S5). For Scipion processing, two-dimensional classification from RELION2.0 was used to remove particles that were either false positives or belonged to classes featuring touching particles, resulting in a data set of 26,624 particles. The selected particles were imported into Scipion (38). Initial models were generated using RANSAC using the obtained two-dimensional averages (40). Three-dimensional classification used the generated initial models after low-pass filtering to 60 Å with C1 symmetry. The selected model was further processed with RELION 3D Auto-Refine in Scipion. For resolution determination, gold-standard Fourier shell correlation (FSC) was used. Back-projections were calculated and compared with the two-dimensional class averages from Xmipp CL2D. For comparison of the SAXS and EM data, EM2DAM (part of the ATSAS software suite) was used to calculate a theoretical X-ray scattering curve from the EM model.

Generation of PcUGGT-biotin mutant expression constructs

The biotin insertion sites were identified following the procedure reported previously (41). We carried out a secondary structure prediction of PcUGGT to identify as many sites as possible to introduce the 15-amino-acid AviTag insertion (LNDILEAQKIEWHEQ) as a biotinylation target across the various domains (Fig. 3*A* and supplemental Fig. S7). A total of 16 sites were identified, presumably loops or unstructured linkers, distributed across the entire sequence of PcUGGT. Primers with AviTag cDNA were synthesized (Biocorp Inc.) and inserted into the pCold I vector bearing the wild-type *PcUGGT* sequence using a QuikChange lightning site-directed mutagenesis Kit (Agilent technologies). The sequences of PcUGGT-biotin mutants were confirmed through sequencing (Genome Quebec).

Preparation of PcUGGT-biotin monovalent streptavidin (mSA)-labeled samples

All the PcUGGT-biotin mutants were expressed in *E. coli* Rosetta Gami 2 as described above for wild-type PcUGGT production. Purification was also done using a similar strategy as the wild-type PcUGGT. Briefly, after the Ni-NTA affinity purification, the protein samples were subjected to *in vitro* biotinylation using the biotin ligase enzyme BirA (41). The successfully biotinylated mutants were then incubated with an excess of monovalent streptavidin (46) and further purified by gel filtration using a Superdex 200 16/600 column.

Negative stain electron microscopy, image processing, and model reconstructions of PcUGGT-biotin-mSA mutants

Preparation of samples for negative-stain EM and data collection was carried out using the same procedures described above for DmUGGT. Briefly, proteins at concentrations of 10 ng/μl were stained with a solution of 0.75% (w/v) uranyl formate and imaged at 62,000× magnification at 200 kV on the FEI Tecnai G2 TF20 transmission electron microscope. We collected roughly 100 micrographs for PcUGGT-biotin-mSA-m3, PcUGGT-biotin-mSA-m7, and PcUGGT-biotin-mSA-m15 separately. Similarly to DmUGGT processing, reference-free classification was applied to sort particles and to select the high-quality particles. The negative-stain PcUGGT model was used as the initial model for refinement of the mSA-labeled mutants. The extra density observed in the mutant maps was attributed to the 50-kDa mSA label. Three-dimensional visualizations of the map were conducted in UCSF Chimera.

Sep15 expression and purification

Human Sep15 with deletion of the first 28 residues (Sep15-Δ28) was cloned into the pET29a vector with a C-terminal hexahistidine tag. The reconstructed plasmid was transformed into lobster *E. coli* strain. Cells were grown until $A_{600\,nm}$ reached 0.6 and then induced with 1 mM isopropyl 1-thio-β-D-galactopyranoside at 30 °C for 4 h. The *E. coli* cells were harvested by centrifugation at 4500 rpm at 4 °C for 15 min and resuspended in Buffer A supplemented with 1 mM DTT. Resuspended cells were lysed using a French pressure cell followed by centrifugation at 12,000 rpm at 4 °C for 1 h. The supernatant

after centrifugation was loaded into the pre-equilibrated Ni-NTA gravity column. After all the supernatant flowed through, the column was washed using Buffer A supplemented with 1 mM DTT. Sep15-Δ28 was eluted from the column using Buffer B supplemented with 1 mM DTT. The elution fraction was concentrated to less than 5 ml using a 10-kDa-molecular-mass-cutoff Amicon filter by centrifugation at 4000 rpm at 4 °C and injected into a Superdex 75 HiLoad 16/600 column, which was pre-equilibrated with Buffer C supplemented with 1 mM DTT. The fractions were analyzed by SDS-PAGE.

Multiangle light scattering analysis

Analytical size-exclusion chromatography (SEC) of DmUGGT and the DmUGGT-Sep15-Δ28 complex was performed using a GE Healthcare 200 10/300GL column with a flow rate of 0.4 ml/min at 22 °C. The SEC was coupled to a mini Dawn TREOS multiangle light scattering analysis instrument (Wyatt Technology) and an Optilab rEX (Wyatt Technology). The measurement of the Rayleigh scattering intensity as a function of the angle as well as the differential refractive index of the eluting peak in SEC can be used to determine the weight average molar mass ($\bar{M}_w$) of eluted peaks, which can be used to estimate the oligomer states and protein complexes using the ASTRA (Wyatt Technologies) (47). The number average molar mass (M_n) was also determined to estimate the monodispersity of the peaks. BSA (4 mg/ml) was used to calibrate the system before running DmUGGT.

Hydrogen-deuterium exchange mass spectrometry

HDX-MS experiments were carried out as described (48). Briefly, stock solutions of 5 mg/ml DmUGGT or the DmUGGT-Sep15 complex were prepared in Buffer C. Similar buffer conditions were applied for the preparation of the corresponding D_2O buffers. For the blank control, the initial dilution was made in H_2O buffer. HDX was initiated by diluting stock DmUGGT solution 1.5:8.5 into the D_2O-based buffer. HDX incubation periods were 15, 300, 900, and 3600 s, and the temperature was set at 25 °C. HDX was quenched with chilled quenching buffer (300 mM glycine, 6 M guanidine hydrochloride, 400 mM tris(2-carboxyethyl)phosphine in H_2O, pH 2.5) using a 1:1 dilution ratio. Quenched samples were flash frozen in methanol containing dry ice, and frozen samples were stored at −80 °C until used.

Prior to UHPLC analysis, deuterated DmUGGT was digested in an on-line immobilized pepsin column prepared in house. Resulting peptides were loaded onto a C_{18} analytical column (1-mm inner diameter, 50 mm; Thermo Fisher Scientific) equipped to an Agilent 1290 UHPLC system. Peptides for each sample were separated using a 5–40% liner gradient of acetonitrile containing 0.1% formic acid for 10 min at a 65 μl/min flow rate. To minimize back-exchange, the column, solvent delivery lines, injector, and other accessories were placed in an ice bath. The C_{18} column was directly connected to the electrospray ionization source of the LTQ Orbitrap XL (Thermo Fisher Scientific), and mass spectra of peptides were acquired in positive-ion mode for m/z 200–2000. The deuteration (%) as a function of incubation time was determined using HDExaminer 2.1 (Sierra Analytics, Modesto, CA). The first two amino acid residues in peptides were excluded from the analysis (49).

Cyanogen bromide-activated Sepharose for affinity binding

Sep15 was coupled to cyanogen bromide-activated Sepharose (GE Healthcare) according to the manufacturer's instructions. The coupling reaction was quenched, and unreacted sites were blocked with 100 mM Tris-HCl buffer, pH 8.0. For pull-down assays, the Sep15-coupled beads were incubated with DmUGGT at room temperature for 1 h in binding buffer (25 mM Tris-HCl, pH 8.0, 300 mM NaCl) and washed three times, and the bound proteins were eluted with SDS-PAGE loading buffer for analysis.

Author contributions—D. C.-G., M. Y., and N. S. collected and analyzed data. M. M., Y. I., and G. K. provided materials. R. M., J. V., J. M. K., and G. L. L. analyzed data. D. C.-G., M. Y., and K. G. wrote the manuscript. All authors agreed on its contents.

Acknowledgments—We thank Dr. Khanh Huy Bui for advice and comments, Drs. Kaustuv Basu and Kelly Sears from the McGill University Facility for Electron Microscopy Research for help in microscope handling and troubleshooting, and Drs. Yoichi Takeda and Akira Seko from the Synthetic Cellular Chemistry Laboratory, RIKEN, for cloning PcUGGT.

References

1. 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
2. Zapun, A., Darby, N. J., Tessier, D. C., Michalak, M., Bergeron, J. J., and Thomas, D. Y. (1998) Enhanced catalysis of ribonuclease B folding by the interaction of calnexin or calreticulin with ERp57. *J. Biol. Chem.* **273,** 6009–6012
3. Kozlov, G., Pocanschi, C. L., Rosenauer, A., Bastos-Aristizabal, S., Gorelik, A., Williams, D. B., and Gehring, K. (2010) Structural basis of carbohydrate recognition by calreticulin. *J. Biol. Chem.* **285,** 38612–38620
4. Kozlov, G., Bastos-Aristizabal, S., Määttänen, P., Rosenauer, A., Zheng, F., Killikelly, A., Trempe, J. F., Thomas, D. Y., and Gehring, K. (2010) Structural basis of cyclophilin B binding by the calnexin/calreticulin P-domain. *J. Biol. Chem.* **285,** 35551–35557
5. Kozlov, G., Maattanen, P., Schrag, J. D., Pollock, S., Cygler, M., Nagar, B., Thomas, D. Y., and Gehring, K. (2006) Crystal structure of the bb′ domains of the protein disulfide isomerase ERp57. *Structure* **14,** 1331–1339
6. D'Alessio, C., Caramelo, J. J., and Parodi, A. J. (2010) UDP-Glc:glycoprotein glucosyltransferase-glucosidase II, the ying-yang of the ER quality control. *Semin. Cell. Dev. Biol.* **21,** 491–499
7. Trombetta, S. E., Gañan, S. A., and Parodi, A. J. (1991) The UDP-Glc: glycoprotein glucosyltransferase is a soluble protein of the endoplasmic reticulum. *Glycobiology* **1,** 155–161
8. Herrero, A. B., Magnelli, P., Mansour, M. K., Levitz, S. M., Bussey, H., and Abeijon, C. (2004) KRE5 gene null mutant strains of *Candida albicans* are avirulent and have altered cell wall composition and hypha formation properties. *Eukaryot. Cell* **3,** 1423–1432
9. Guerin, M., and Parodi, A. J. (2003) The UDP-glucose:glycoprotein glucosyltransferase is organized in at least two tightly bound domains from yeast to mammals. *J. Biol. Chem.* **278,** 20540–20546
10. Choudhury, P., Liu, Y., Bick, R. J., and Sifers, R. N. (1997) Intracellular association between UDP-glucose:glycoprotein glucosyltransferase and an incompletely folded variant of α_1-antitrypsin. *J. Biol. Chem.* **272,** 13446–13451
11. Tessier, D. C., Dignard, D., Zapun, A., Radominska-Pandya, A., Parodi, A. J., Bergeron, J. J., and Thomas, D. Y. (2000) Cloning and characteriza-

tion of mammalian UDP-glucose glycoprotein:glucosyltransferase and the development of a specific substrate for this enzyme. *Glycobiology* **10,** 403–412
12. Tannous, A., Patel, N., Tamura, T., and Hebert, D. N. (2015) Reglucosylation by UDP-glucose:glycoprotein glucosyltransferase 1 delays glycoprotein secretion but not degradation. *Mol. Biol. Cell* **26,** 390–405
13. Ritter, C., Quirin, K., Kowarik, M., and Helenius, A. (2005) Minor folding defects trigger local modification of glycoproteins by the ER folding sensor GT. *EMBO J.* **24,** 1730–1738
14. Taylor, S. C., Thibault, P., Tessier, D. C., Bergeron, J. J., and Thomas, D. Y. (2003) Glycopeptide specificity of the secretory protein folding sensor UDP-glucose glycoprotein:glucosyltransferase. *EMBO Rep.* **4,** 405–411
15. Izumi, M., Makimura, Y., Dedola, S., Seko, A., Kanamori, A., Sakono, M., Ito, Y., and Kajihara, Y. (2012) Chemical synthesis of intentionally misfolded homogeneous glycoprotein: a unique approach for the study of glycoprotein quality control. *J. Am. Chem. Soc.* **134,** 7238–7241
16. Dedola, S., Izumi, M., Makimura, Y., Seko, A., Kanamori, A., Sakono, M., Ito, Y., and Kajihara, Y. (2014) Folding of synthetic homogeneous glycoproteins in the presence of a glycoprotein folding sensor enzyme. *Angew. Chem. Int. Ed. Engl.* **53,** 2883–2887
17. Totani, K., Ihara, Y., Matsuo, I., Koshino, H., and Ito, Y. (2005) Synthetic substrates for an endoplasmic reticulum protein-folding sensor, UDP-glucose:glycoprotein glucosyltransferase. *Angew. Chem. Int. Ed. Engl.* **44,** 7950–7954
18. Takeda, Y., Totani, K., Matsuo, I., and Ito, Y. (2009) Chemical approaches toward understanding glycan-mediated protein quality control. *Curr. Opin. Chem. Biol.* **13,** 582–591
19. Totani, K., Ihara, Y., Tsujimoto, T., Matsuo, I., and Ito, Y. (2009) The recognition motif of the glycoprotein-folding sensor enzyme UDP-Glc: glycoprotein glucosyltransferase. *Biochemistry* **48,** 2933–2940
20. Sakono, M., Seko, A., Takeda, Y., Hachisu, M., and Ito, Y. (2012) Biophysical properties of UDP-glucose:glycoprotein glucosyltransferase, a folding sensor enzyme in the ER, delineated by synthetic probes. *Biochem. Biophys. Res. Commun.* **426,** 504–510
21. Takeda, Y., Seko, A., Hachisu, M., Daikoku, S., Izumi, M., Koizumi, A., Fujikawa, K., Kajihara, Y., and Ito, Y. (2014) Both isoforms of human UDP-glucose:glycoprotein glucosyltransferase are enzymatically active. *Glycobiology* **24,** 344–350
22. Ohara, K., Takeda, Y., Daikoku, S., Hachisu, M., Seko, A., and Ito, Y. (2015) Profiling aglycon-recognizing sites of UDP-glucose:glycoprotein glucosyltransferase by means of squarate-mediated labeling. *Biochemistry* **54,** 4909–4917
23. Taylor, S. C., Ferguson, A. D., Bergeron, J. J., and Thomas, D. Y. (2004) The ER protein folding sensor UDP-glucose glycoprotein-glucosyltransferase modifies substrates distant to local changes in glycoprotein conformation. *Nat. Struct. Mol. Biol.* **11,** 128–134
24. Ritter, C., and Helenius, A. (2000) Recognition of local glycoprotein misfolding by the ER folding sensor UDP-glucose:glycoprotein glucosyltransferase. *Nat. Struct. Biol.* **7,** 278–280
25. Korotkov, K. V., Kumaraswamy, E., Zhou, Y., Hatfield, D. L., and Gladyshev, V. N. (2001) Association between the 15-kDa selenoprotein and UDP-glucose:glycoprotein glucosyltransferase in the endoplasmic reticulum of mammalian cells. *J. Biol. Chem.* **276,** 15330–15336
26. Ferguson, A. D., Labunskyy, V. M., Fomenko, D. E., Araç, D., Chelliah, Y., Amezcua, C. A., Rizo, J., Gladyshev, V. N., and Deisenhofer, J. (2006) NMR structures of the selenoproteins Sep15 and SelM reveal redox activity of a new thioredoxin-like family. *J. Biol. Chem.* **281,** 3536–3543
27. Labunskyy, V. M., Ferguson, A. D., Fomenko, D. E., Chelliah, Y., Hatfield, D. L., and Gladyshev, V. N. (2005) A novel cysteine-rich domain of Sep15 mediates the interaction with UDP-glucose:glycoprotein glucosyltransferase. *J. Biol. Chem.* **280,** 37839–37845
28. Wearsch, P. A., Peaper, D. R., and Cresswell, P. (2011) Essential glycan-dependent interactions optimize MHC class I peptide loading. *Proc. Natl. Acad. Sci. U.S.A.* **108,** 4950–4955
29. Zhang, W., Wearsch, P. A., Zhu, Y., Leonhardt, R. M., and Cresswell, P. (2011) A role for UDP-glucose glycoprotein glucosyltransferase in expression and quality control of MHC class I molecules. *Proc. Natl. Acad. Sci. U.S.A.* **108,** 4956–4961
30. Arnold, S. M., and Kaufman, R. J. (2003) The noncatalytic portion of human UDP-glucose:glycoprotein glucosyltransferase I confers UDP-glucose binding and transferase function to the catalytic domain. *J. Biol. Chem.* **278,** 43320–43328
31. Zhu, T., Satoh, T., and Kato, K. (2014) Structural insight into substrate recognition by the endoplasmic reticulum folding-sensor enzyme: crystal structure of third thioredoxin-like domain of UDP-glucose:glycoprotein glucosyltransferase. *Sci. Rep.* **4,** 7322
32. Arnold, S. M., Fessler, L. I., Fessler, J. H., and Kaufman, R. J. (2000) Two homologues encoding human UDP-glucose:glycoprotein glucosyltransferase differ in mRNA expression and enzymatic activity. *Biochemistry* **39,** 2149–2163
33. Parker, C. G., Fessler, L. I., Nelson, R. E., and Fessler, J. H. (1995) *Drosophila* UDP-glucose:glycoprotein glucosyltransferase: sequence and characterization of an enzyme that distinguishes between denatured and native proteins. *EMBO J.* **14,** 1294–1303
34. Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M., Gorba, C., Mertens, H. D., Konarev, P. V., and Svergun, D. I. (2012) New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **45,** 342–350
35. Baldwin, P. R., and Penczek, P. A. (2007) The transform class in SPARX and EMAN2. *J. Struct. Biol.* **157,** 250–261
36. Tang, G., Peng, L., Baldwin, P. R., Mann, D. S., Jiang, W., Rees, I., and Ludtke, S. J. (2007) EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157,** 38–46
37. Kimanius, D., Forsberg, B. O., Scheres, S. H., and Lindahl, E. (2016) Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *eLife* **5,** e18722
38. de la Rosa-Trevín, J. M., Quintana, A., Del Cano, L., Zaldívar, A., Foche, I., Gutiérrez, J., Gómez-Blanco, J., Burguet-Castell, J., Cuenca-Alba, J., Abrishami, V., Vargas, J., Otón, J., Sharov, G., Vilas, J. L., Navas, J., *et al.* (2016) Scipion: a software framework toward integration, reproducibility and validation in 3D electron microscopy. *J. Struct. Biol.* **195,** 93–99
39. Sorzano, C. O., Bilbao-Castro, J. R., Shkolnisky, Y., Alcorlo, M., Melero, R., Caffarena-Fernández, G., Li, M., Xu, G., Marabini, R., and Carazo, J. M. (2010) A clustering approach to multireference alignment of single-particle projections in electron microscopy. *J. Struct. Biol.* **171,** 197–206
40. Vargas, J., Álvarez-Cabrera, A. L., Marabini, R., Carazo, J. M., and Sorzano, C. O. (2014) Efficient initial volume determination from electron microscopy images of single particles. *Bioinformatics* **30,** 2891–2898
41. Lau, P. W., Potter, C. S., Carragher, B., and MacRae, I. J. (2012) DOLORS: versatile strategy for internal labeling and domain localization in electron microscopy. *Structure* **20,** 1995–2002
42. Sousa, M. C., Ferrero-Garcia, M. A., and Parodi, A. J. (1992) Recognition of the oligosaccharide and protein moieties of glycoproteins by the UDP-Glc:glycoprotein glucosyltransferase. *Biochemistry* **31,** 97–105
43. Caramelo, J. J., Castro, O. A., de Prat-Gay, G., and Parodi, A. J. (2004) The endoplasmic reticulum glucosyltransferase recognizes nearly native glycoprotein folding intermediates. *J. Biol. Chem.* **279,** 46280–46285
44. Elmlund, D., and Elmlund, H. (2012) SIMPLE: software for *ab initio* reconstruction of heterogeneous single-particles. *J. Struct. Biol.* **180,** 420–427
45. Elmlund, H., Elmlund, D., and Bengio, S. (2013) PRIME: probabilistic initial 3D model generation for single-particle cryo-electron microscopy. *Structure* **21,** 1299–1306
46. Howarth, M., Chinnapen, D. J., Gerrow, K., Dorrestein, P. C., Grandy, M. R., Kelleher, N. L., El-Husseini, A., and Ting, A. Y. (2006) A monovalent streptavidin with a single femtomolar biotin binding site. *Nat. Methods* **3,** 267–273
47. Trathnigg, B. (1995) Determination of MWD and chemical composition of polymers by chromatographic techniques. *Prog. Polym. Sci.* **20,** 615–650
48. Okiyoneda, T., Veit, G., Dekkers, J. F., Bagdany, M., Soya, N., Xu, H., Roldan, A., Verkman, A. S., Kurth, M., Simon, A., Hegedus, T., Beekman, J. M., and Lukacs, G. L. (2013) Mechanism-based corrector combination restores ΔF508-CFTR folding and function. *Nat. Chem. Biol.* **9,** 444–454
49. Bai, Y., Milne, J. S., Mayne, L., and Englander, S. W. (1993) Primary structure effects on peptide group hydrogen exchange. *Proteins* **17,** 75–86