Interdomain conformational flexibility underpins the activity of UGGT, the eukaryotic glycoprotein secretion checkpoint

Pietro Roversi[a,1,2], Lucia Marti[b,1], Alessandro T. Caputo[a,1], Dominic S. Alonzi[a,1], Johan C. Hill[a,1], Kyle C. Dent[c], Abhinav Kumar[a], Mikail D. Levasseur[a], Andrea Lia[a,b], Thomas Waksman[a], Souradeep Basu[a], Yentli Soto Albrecht[a], Kristin Qian[a], James Patrick McIvor[a], Colette B. Lipp[a], Dritan Siliqi[d], Snežana Vasiljević[a], Shabaz Mohammed[a], Petra Lukacik[c], Martin A. Walsh[c], Angelo Santino[b], and Nicole Zitzmann[a,2]

[a]Oxford Glycobiology Institute, Department of Biochemistry, University of Oxford, Oxford OX1 3QU, United Kingdom; [b]Institute of Sciences of Food Production, Consiglio Nazionale delle Ricerche (CNR) Unit of Lecce, I-73100 Lecce, Italy; [c]Diamond Light Source and Research Complex at Harwell, Didcot OX11 0DE, United Kingdom; and [d]Istituto di Cristallografia, Consiglio Nazionale delle Ricerche, I-70126 Bari, Italy

Edited by Peter Cresswell, Yale University School of Medicine, New Haven, CT, and approved June 26, 2017 (received for review March 4, 2017)

Glycoproteins traversing the eukaryotic secretory pathway begin life in the endoplasmic reticulum (ER), where their folding is surveyed by the 170-kDa UDP-glucose:glycoprotein glucosyltransferase (UGGT). The enzyme acts as the single glycoprotein folding quality control checkpoint: it selectively reglucosylates misfolded glycoproteins, promotes their association with ER lectins and associated chaperones, and prevents premature secretion from the ER. UGGT has long resisted structural determination and sequence-based domain boundary prediction. Questions remain on how this single enzyme can flag misfolded glycoproteins of different sizes and shapes for ER retention and how it can span variable distances between the site of misfold and a glucose-accepting N-linked glycan on the same glycoprotein. Here, crystal structures of a full-length eukaryotic UGGT reveal four thioredoxin-like (TRXL) domains arranged in a long arc that terminates in two β-sandwiches tightly clasping the glucosyltransferase domain. The fold of the molecule is topologically complex, with the first β-sandwich and the fourth TRXL domain being encoded by nonconsecutive stretches of sequence. In addition to the crystal structures, a 15-Å cryo-EM reconstruction reveals interdomain flexibility of the TRXL domains. Double cysteine point mutants that engineer extra interdomain disulfide bridges rigidify the UGGT structure and exhibit impaired activity. The intrinsic flexibility of the TRXL domains of UGGT may therefore endow the enzyme with the promiscuity needed to recognize and reglucosylate its many different substrates and/or enable reglucosylation of N-linked glycans situated at variable distances from the site of misfold.

UGGT | endoplasmic reticulum | glycoprotein folding | UDP-glucose glycoprotein glucosyltransferase | eukaryotic secretion

About one-third of eukaryotic genomes code for proteins that are destined for the secretory pathway, and of these, around 70% are *N*-glycosylated (1). They emerge from the ribosomes into the endoplasmic reticulum (ER) lumen in an unfolded state (2), and their folding progress is monitored by the UDP-glucose:glycoprotein glucosyltransferase (UGGT) (3), a 170-kDa ER-resident enzyme that selectively recognizes and reglucosylates only misfolded glycoproteins (4, 5) and misassembled glycoprotein complexes (6). The glucose molecule is transferred by UGGT from UDP-glucose to an N-linked glycan on the misfolded glycoprotein (7, 8). A glycoprotein bearing a monoglucosylated N-linked glycan is retained in the ER bound to the lectins calnexin and/or calreticulin and the associated chaperones and foldases that assist folding (9). Prolonged UGGT-mediated ER retention ultimately leads to ER-associated degradation (ERAD) (10).

Most vertebrates have two homologous genes for UGGT, sharing a 55% sequence identity: UGGT1 and UGGT2. The former binds UDP-Glc and reglucosylates misfolded glycoproteins. UGGT1 is essential during early organism development and homozygous UGGT1$^{-/-}$ knockout of the gene is embryonically lethal in mice, although cells derived from those embryos are viable (11). More recently, heterozygous UGGT1$^{+/-}$ knockout mice have been reported to express approximately half of the wild-type (WT) amount of UGGT1 but they undergo normal development and have no obvious aberrant phenotype (12).

In healthy cells enjoying steady-state glycoprotein homeostasis, glycoproteins that are slow or difficult to fold need repeated cycles of association with the ER lectins and chaperones. To this effect, they undergo multiple UGGT-mediated reglucosylation cycles (2, 13–15). UGGT expression is increased upon ER stress and plays an important role in the unfolded protein response (16). The enzyme also surveys the assembly of key immunological molecules, the T-cell receptor (TCR) and the major histocompatibility complex (MHC). Four of the six TCR subunits carry an N-linked glycan, and UGGT continues to reglucosylate them until proper disulfide linkages are established and the whole TCR complex assembly is complete (6). MHC class I molecules that fail to load a peptide (17) or are associated with a suboptimal one (18) are also preferentially recognized and

Significance

A dedicated endoplasmic reticulum quality control (ERQC) machinery ensures the correct fold of secreted proteins bearing N-linked glycans, which constitute around a fifth of the whole proteome and are essential for many important cellular processes such as signaling, immunity, adhesion, transport, and metabolism. UDP-glucose:glycoprotein glucosyltransferase (UGGT) is the sole checkpoint enzyme of ERQC, flagging incorrectly folded glycoproteins for ER retention. Here, we describe crystal structures of full-length UGGT. We show that enzymatic activity depends on interdomain conformational mobility, indicating that the intrinsic flexibility of UGGT may endow the enzyme with the promiscuity needed to recognize and reglucosylate its many different substrates.

Author contributions: P.R., L.M., A.T.C., D.S.A., J.C.H., A.S., and N.Z. designed research; P.R., L.M., A.T.C., D.S.A., J.C.H., K.C.D., A.K., M.D.L., A.L., T.W., S.B., Y.S.A., K.Q., J.P.M., C.B.L., D.S., S.V., S.M., P.L., M.A.W., and A.S. performed research; P.R., L.M., A.T.C., D.S.A., J.C.H., K.C.D., A.K., M.D.L., A.L., T.W., S.B., Y.S.A., K.Q., J.P.M., C.B.L., D.S., S.V., S.M., P.L., M.A.W., A.S., and N.Z. analyzed data; and P.R., A.T.C., D.S.A., J.C.H., and N.Z. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 5N2J, 5MU1, 5MZO, and 5NV4).

[1]P.R., L.M., A.T.C., D.S.A., and J.C.H. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: pietro.roversi@bioch.ox.ac.uk or nicole.zitzmann@bioch.ox.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1703682114/-/DCSupplemental.

reglucosylated by UGGT, leading to their ER retention and adding an extra level of control to MHC I antigen selection and presentation. Whereas UGGT activity is beneficial to avoid premature secretion of healthy glycoproteins under physiological conditions, in the background of genetic mutations that impair the fold but not the function of a glycoprotein ("responsive mutants") (19) UGGT-mediated ER retention exacerbates the consequences of minor folding defects (19–22), causing disease. For example, UGGT interacts with the cystic fibrosis transmembrane conductance regulator ΔF508 mutant (CFTR-ΔF508) responsible for 70% of cystic fibrosis cases (22). No UGGT inhibitors (other than UDP) (23) are known, so the extent to which partial inhibition of UGGT can ameliorate congenital protein misfolding disease remains to be tested (3, 20, 24).

The N-terminal ∼1,200 residues of UGGT harbor the enzyme's misfold sensing activity, whereas the C-terminal ∼300 residues encode a glucosyltransferase 24 family (GT24) A-type domain (25–27). Full-length UGGT has long resisted structural determination (3) and sequence-based domain boundary prediction can only reliably detect three domains, of which thioredoxin-like domain 3 (TRXL3) is the only UGGT domain for which a structure is available (28). The mechanism by which UGGT recognizes and reglucosylates its substrates, which differ considerably in size and shape (11, 22, 29–32), remains unknown. The enzyme also needs to span variable distances between the site of misfold and the glucose-accepting N-linked glycan on the same glycoprotein (33–35). To aid our understanding of this pivotal molecular sensor device at the heart of every cell's glycoprotein folding machinery, we cloned, expressed, and purified UGGT from the thermophilic yeast *Chaetomium thermophilum* (36) (*Ct*UGGT) and characterized it structurally and biochemically.

Results and Discussion

UGGT Has a Novel Seven-Domain Fold of Complex Topology. Four different crystal forms of *Ct*UGGT (Fig. 1 and *SI Appendix*, Fig. S1 *A* and *B* and Tables S1–S3) contain five crystallographically independent molecules of 6 × 8 × 12 nm dimensions, organized in seven domains (Fig. 1 *A*–*C* and *SI Appendix*, Fig. S2). In the N-terminal region, four thioredoxin-like domains (TRXL1–4) form an extended arc, capped at one end by two seven-stranded β-sandwiches (βS1 and βS2) that tightly clasp the C-terminal catalytic glucosyltransferase (GT) domain (Fig. 1 *A*–*C*). Only the TRXL2, TRXL3, and GT domain boundaries were correctly predicted from sequence (28, 37). Unusually, the TRXL1 domain folds with sequential pairing of a four-helix subdomain (*Ct*UGGT residues 43–110) with a thioredoxin subdomain (*Ct*UGGT residues 111–216), giving UGGT–TRXL1 secondary structure: ααααα–βαβαββα. All other known TRXL domain structures present the four-helix subdomain as an insertion within the thioredoxin subdomain (canonical TRXL secondary structure: βαβ–ααααα–αββα). Even more unusually, the TRXL4 and βS1 subdomains are encoded by nonconsecutive stretches of sequence (Fig. 1 *D* and *E*). The complex topology of UGGT caused it to escape all previous attempts at sequence-based subdomain boundary and fold prediction and raises the question how much chaperoning the ER glycoprotein misfold sensor itself may need to fold correctly.

Catalytic and Misfold-Recognition Domains Are Situated at Opposite Ends of the Molecule. Across UGGT sequences in higher eukaryotes, the GT catalytic domain surface shows areas of high conservation, likely mapping to the binding regions for UDP-Glc and the substrate glycans (*SI Appendix*, Figs. S3 and S4*A*). The GT domain displays the expected glycosyltransferase type A (GT-A) fold (*SI Appendix*, Fig. S4*B*) (27). *SI Appendix*, Fig. S4*C* illustrates

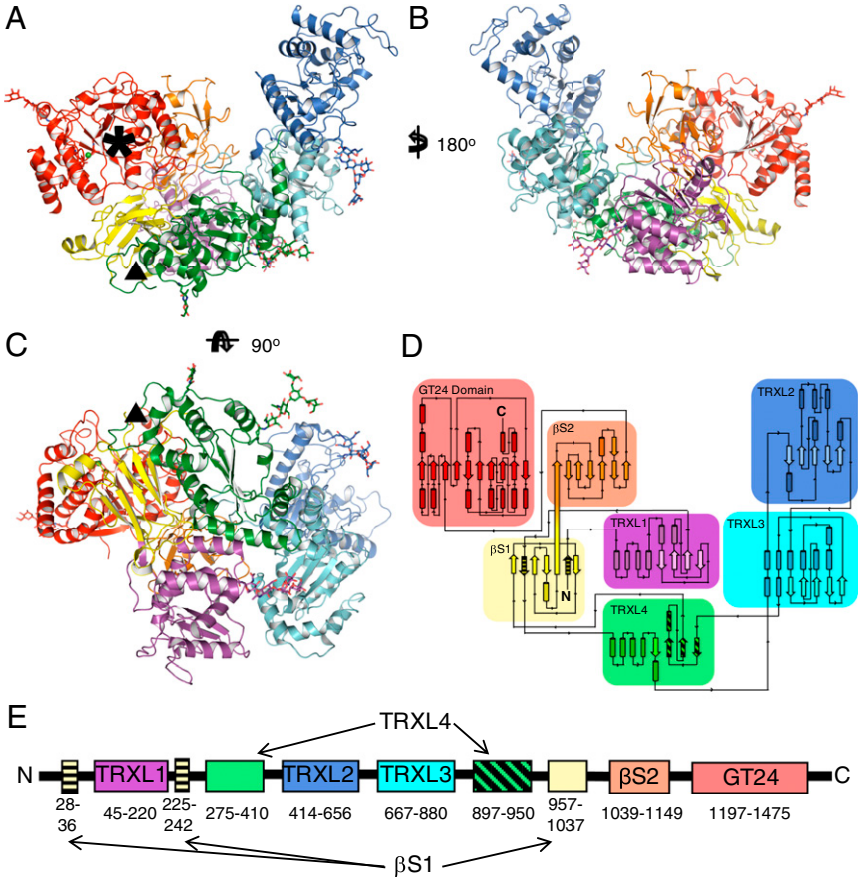

Fig. 1. Structure and topology of *Ct*UGGT. (*A*–*C*) Three orthogonal views of the 3.5-Å crystal structure of the *Ct*UGGT $P6_1$ (intermediate) form. The structure spans residues 27–1,473, except for disordered loops 243–285 and 1,334–1,340, and the residue gap 1,153–1,195 around the endoproteolysis site, also disordered in the crystal. The seven domains are in cartoon representation: TRXL1 (purple), TRXL2 (blue), TRXL3 (cyan), TRXL4 (green), βS1 (yellow), βS2 (orange), and GT (red). The five N-linked glycans (at N56, N329, N638, N894, and N1227) are in stick representation. Six conserved cysteine residues form three disulfide bonds, one in the TRXL1 domain (*Ct*UGGT C138–C150) and two in the GT domain (*Ct*UGGT C1330–C1423 and C1419–C1437). The Ca^{2+} ion bound at the conserved nucleotide-sugar coordinating ^{1302}DAD^{1304} motif is represented as a green sphere. A black asterisk marks the catalytic site. A black triangle marks the dangling ends around the disordered loop (*Ct*UGGT residues 243–285, between the second strand of the βS1 sandwich and the N-terminal part of the TRXL4 domain) corresponding to the region to which Sep15 binding has been mapped in *D. melanogaster* UGGT (37). (*D* and *E*) The 2- and 1-dimensional topological diagrams of *Ct*UGGT, respectively. The first and second strands of β-sandwich βS1 (residues 28–36 and 225–242, respectively, striped yellow) flank the part of sequence encoding TRXL1 (residues 45–220, magenta), with the rest of βS1 encoded by a portion of sequence (residues 957–1,037, yellow) more than 700 residues downstream of the sandwich's second strand. TRXL1 adopts a noncanonical subdomain structure, different from all structures of Pfam family PF01323 members, in which an N-terminal α-helical subdomain is followed by a C-terminal thioredoxin subdomain. The N- and C-terminal halves of TRXL4 (residues 275–410 and 897–950, the latter portion in striped green) occur in the standard order but are separated in sequence by TRXL2 and TRXL3.

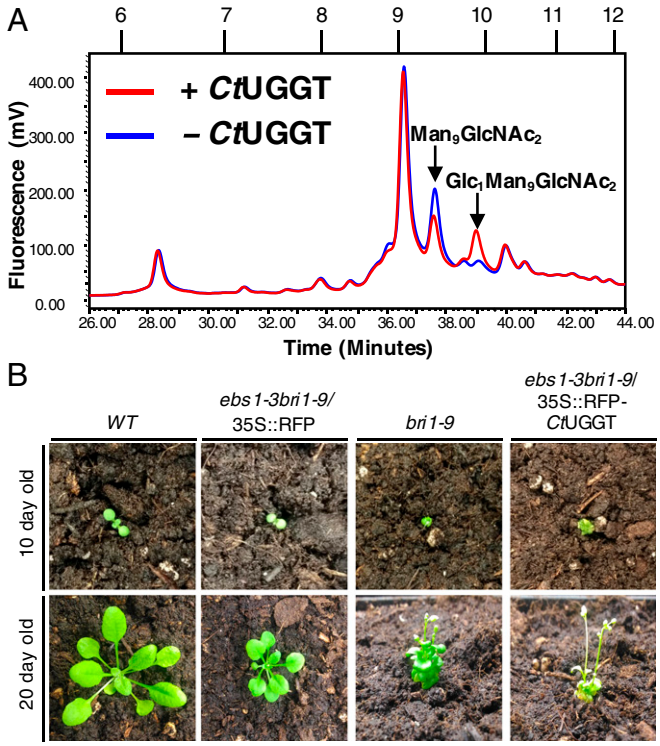

Fig. 2. *Ct*UGGT is active in vitro at 37 °C and in vivo at 25 °C. (*A*) In vitro *Ct*UGGT-mediated reglucosylation of urea-misfolded bovine thyroglobulin at 37 °C. The HPLC elution profile of the N-linked glycans obtained from urea-misfolded bovine thyroglobulin is in blue. In red, the glycans from the same protein after treatment with *Ct*UGGT at 37 °C. The amount of $Man_9GlcNAc_2$ is reduced and the amount of $Glc_1Man_9GlcNAc_2$ increases, proving that the enzyme is active. (*B*) *Ct*UGGT rescues the *bri1-9* phenotype. Representative images of double mutant plants *ebs1-3 bri1-9* stably expressing the red-fluorescent protein *Ct*UGGT fusion 35S::RFP–*Ct*UGGT are compared with WT, *ebs1-3 bri1-9*, and *bri1-9* plants at 10 and 20 d after germination. The typical dwarf phenotype of the *bri1-9* mutant was restored in the *ebs1-3 bri1-9* genetic background (which carries an inactive *At*UGGT gene and therefore displays a phenotype similar to WT) by the functional *Ct*UGGT protein.

putative UDP-Glc coordinating UGGT residues, based on the structural homology with glycosyltransferases crystallized with nucleotide sugar donors (38).

At the other end of the molecule, past the two β-sandwiches, the presence of TRXL domains in the N-terminal portion of UGGT is reminiscent of domains of the same fold belonging to ER luminal chaperones such as *Drosophila melanogaster* Wind (39, 40), mammalian ERp29 and ERp46 (41, 42), and ER protein disulfide isomerases (43, 44). All these proteins have the ability of binding misfolded proteins and/or peptides. In higher eukaryotes with a UGGT1 and a UGGT2 gene, only the amino-terminal portion of UGGT1, and not the equivalent portion of UGGT2, can recognize misfolded proteins (26, 45). Therefore, differences in conserved surface residues between UGGT1 and UGGT2 inform as to potential misfold recognition hotspots. In particular, the surfaces of the TRXL2 and TRXL3 domains feature residues conserved across UGGT1 sequences but not across UGGT2 sequences (boxed in *SI Appendix*, Fig. S3 *A* and *B*). Interestingly, these UGGT1 conserved residues are located on the same face of the molecule as the large conserved patch around the active site. Further work will be needed to explore the significance of such observations and pinpoint both the UGGT misfold recognition regions and its molecular mechanism.

***Ct*UGGT Is Active in Vitro at 37 °C and in Vivo at 25 °C.** UGGT protein and function are widely conserved across eukaryotes (3), the only known exceptions being a few protists and fungi that make either extremely short N-linked glycans or no N-linked glycans at all (46, 47). We set out to test *Ct*UGGT activity in vitro and in vivo. Circular dichroism (CD) spectroscopy in the temperature range 20–60 °C detects no major changes to the enzyme secondary structure (*SI Appendix*, Fig. S1 *E* and *F*). *Ct*UGGT is active in vitro at 37 °C in reglucosylating urea-denatured bovine thyroglobulin, a well-known UGGT substrate (5, 33) and likely an ensemble of misfolded molecules (48) (Fig. 2*A*). The cellular localization of *Ct*UGGT in vivo at 25 °C (49) was assayed by transfecting the leaves of tobacco and *Arabidopsis thaliana* (*At*) plants with a red-fluorescent protein fusion of *Ct*UGGT (35S::RFP–CtUGGT): fluorescence microscopy confirms that the RFP–*Ct*UGGT fusion protein correctly localizes to the plant ER (*SI Appendix*, Fig. S5 *A* and *B*).

To check the in vivo activity of the protein, we chose genetic recomplementation of *Ct*UGGT in the well-studied *A. thaliana* double mutant *At ebs1-3 bri1-9* plant, the first and only example of a living organism whose growth phenotype depends on UGGT activity (30). The double mutant *At ebs1-3 bri1-9* plant is derived from the *At bri1-9* strain, which carries a brassinosteroid hormone receptor gene *bri1* point mutation that impairs the fold of the BRI1 receptor but does not abrogate its activity (mutation *bri1-9*) (50). The *bri1-9* point mutation leads to

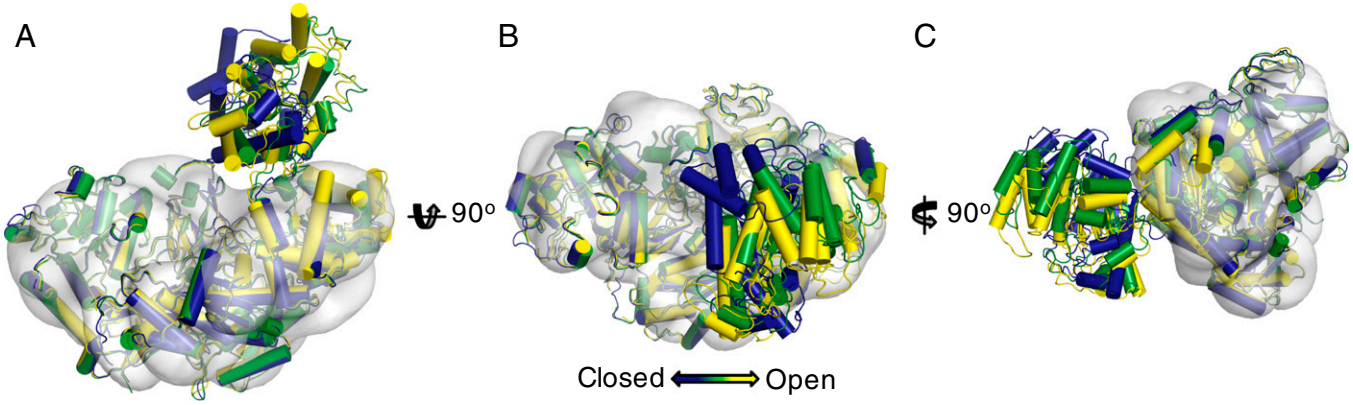

Fig. 3. The 15.2-Å cryo-EM *Ct*UGGT reconstruction confirms mobility of TRXL2. (*A*–*C*) Three orthogonal views of the overlay of three crystallographically independent molecules of WT *Ct*UGGT across three crystal forms. Helices are represented by cylinders. Yellow: $P6_122$ form, open conformation; green: $P6_1$ form, intermediate conformation; and blue: $P4_3$ form, closed conformation (molecule A). The models have been superimposed with the main class of the 15.2-Å cryo-EM reconstruction. The cryo-EM reconstruction contour level encloses a volume corresponding to the protein mass.

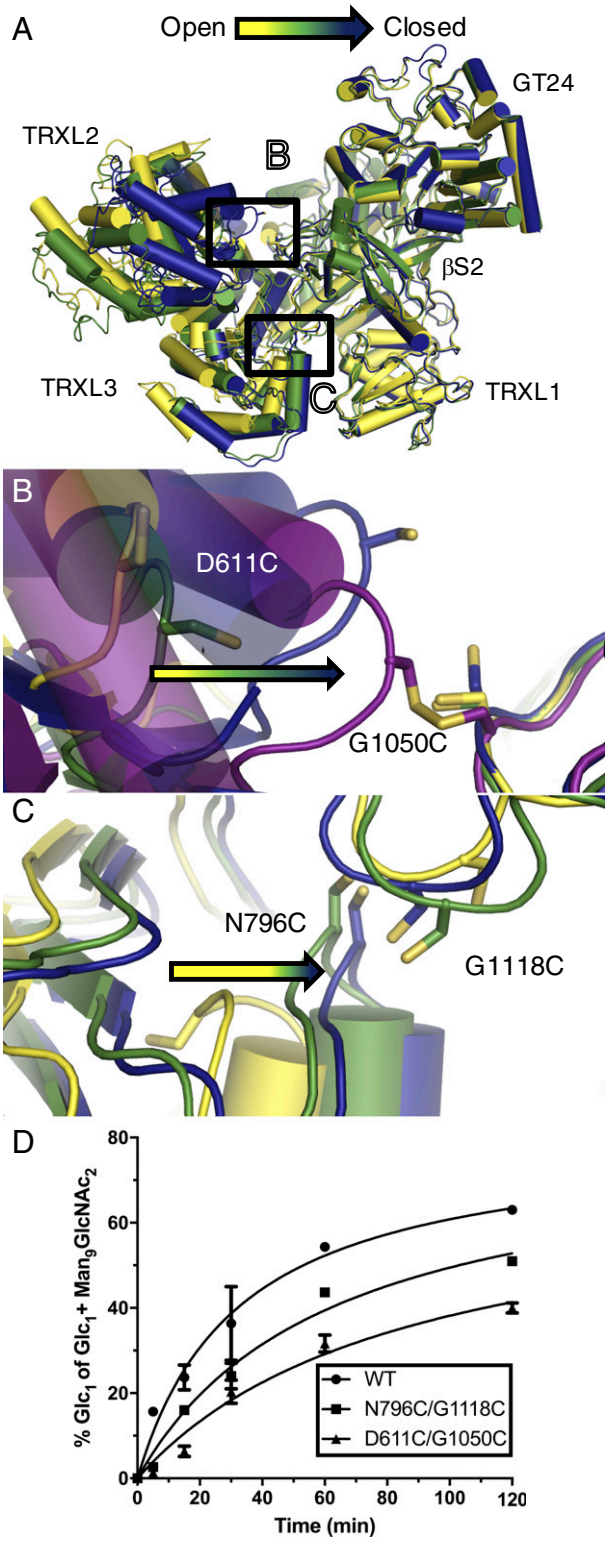

Fig. 4. UGGT interdomain conformational mobility underpins its activity. (*A*) Overlay of three crystallographically independent molecules of WT *Ct*UGGT across three crystal forms, colored as in Fig. 3. (*B*) Zoom into the interface between the TRXL2 and βS2 domains, with the Cys residues introduced by the mutations D611C and G1050C modeled in stick representation for the WT structures, and the observed disulfide bridge in the structure of the $CtUGGT^{D611C/G1050C}$ double mutant in magenta. The double mutant $CtUGGT^{D611C/G1050C}$ prevents *Ct*UGGT from acquiring the open/intermediate conformations. (*C*) Zoom into the interface between the TRXL3 and βS2 domains, with the Cys residues introduced by the mutations N796C and G1118C modeled in stick representation. The double mutant $CtUGGT^{N796C/G1118C}$ locks *Ct*UGGT away from the open conformation. (*D*) Time course of the reglucosylation of urea-denatured bovine thyroglobulin mediated by wild-type (WT, circles) and double Cys mutants of *Ct*UGGT ($CtUGGT^{N796C/G1118C}$, squares; $CtUGGT^{D611C/G1050C}$, triangles). Each data point comes from three independent reglucosylation experiments. The amount of UGGT-mediated reglucosylation is determined as the ratio of the amount of $Glc_1Man_9GlcNAc_2$ to the amount of ($Glc_1Man_9GlcNAc_2$ + $Man_9GlcNAc_2$). The double Cys mutants are less active than the wild-type enzyme.

UGGT-mediated ER retention of the BRI1-9 receptor and consequently a dwarf phenotype. In the *At ebs1-3 bri1-9* strain, which carries an inactive *At* UGGT gene (mutation *ebs1-3*) as well as the *bri1-9* point mutation, a phenotype similar to wild type is observed: in the absence of UGGT, the misfolded and yet active BRI1-9 receptor is not retained in the ER; it is secreted and can signal for plant growth (30).

Upon transfection of the *At ebs1-3 bri1-9* double mutant plants with a RFP–*Ct*UGGT fusion protein, the enzyme correctly localizes to the ER (*SI Appendix*, Fig. S5 *B* and *C*) and these plants revert to the *bri1-9* dwarf phenotype (Fig. 2*B*), confirming that *Ct*UGGT is active in reglucosylating the misfolded BRI1-9 receptor in vivo at 25 °C.

GT and βS1/βS2 Domains Form a Rigid Substructure Spanning the UGGT Cleavable Flexible Linker. In keeping with observations with rat, fruitfly, and *Schizosaccharomyces pombe* UGGTs (25, 51), the N- and C-terminal parts of *Ct*UGGT are cleaved by endoproteolysis at a site in the flexible linker connecting the βS2 domain to the GT catalytic domain (*SI Appendix*, Fig. S1*B*). EDTA treatment suggests that the endoproteolysis is divalent metal dependent (*SI Appendix*, Fig. S1*C*). N-terminal sequencing of recombinant *Ct*UGGT confirms that the endoproteolysis takes place around the stretch *Ct*UGGT 1166–1175. The flexible linker containing the endoproteolytically sensitive portion between the catalytic and folding sensor domains of UGGT was speculated to enable the spanning of both long and short distances between the site of misfold and the glucose-accepting glycan (34). In the crystal structures, no ordered electron density is visible on either side of the endoproteolysis site, past *Ct*UGGT βS2 residue P1152 and before *Ct*UGGT GT residue E1196. On either side of the endoproteolysis site, the ∼1,500 $Å^2$ interface between the β-sandwiches and the GT domain is structurally well conserved across the four WT crystallographically independent molecules (overall C_α rmsd = 0.79 Å for 514 residues in the GT, βS1, and βS2 domains). These observations are consistent with the N- and C-terminal UGGT endoproteolytic fragments being tightly and stably associated in solution (*SI Appendix*, Fig. S1*D*) (25, 51). In other words, the region between the βS2 and GT domains is indeed flexible, but it is not allowing relative movement of the portions of structure immediately preceding and following it, dismissing the hypothesis of the flexible linker in this region as the main source of the protein's versatility.

UGGT TRXL2 and TRXL3 Domains Are Flexible. Our crystal structures show that the most mobile UGGT domain is TRXL2, which is loosely attached to the rest of the protein via hinge points at its N and C termini (*Ct*UGGT residues 415–418 and 651–654). In the three crystal structures, TRXL2 adopts three different orientations, which we term "open" (space group $P6_122$), "intermediate" (space group $P6_1$), and "closed" (space group $P4_3$) conformations. Moving from the open to the intermediate conformation, the TRXL2 and TRXL3 domains rotate by 8° and 10° and move closer to the main body of the protein by about 3 and 5 Å, respectively (Movies S1*A* and S1*B*). Much more pronounced is the transition between the intermediate and the closed forms: TRXL3 swings slightly away from βS2 and TRXL2 rotates in a further 40°, coming even closer to βS2 (Movies S1*A* and S1*B*). Small angle X-ray scattering (SAXS) of *Ct*UGGT detects only limited conformational mobility (*SI Appendix*, Table S4 and Fig. S6), so it is likely that the limited

range of protein conformations observed in the crystals is representative of those in solution. Conformational mobility of TRXL2 is further supported by cryoelectron microscopy of *Ct*UGGT: the sorting of about 64,000 *Ct*UGGT particle views detects molecules belonging to four different types ("classes") (Fig. 3 and *SI Appendix*, Fig. S7). The main class is described by a 15.2-Å cryo-EM reconstruction that agrees with the X-ray structures in the main body of the molecule [correlation coefficient (CC) = 0.83] but lacks defined density for TRXL2 (Fig. 3). Masking to exclude TRXL2 was necessary for 3D reconstruction (*SI Appendix*, Fig. S7), suggesting different relative orientations between this domain and the rest of the protein. High mobility within the *D. melanogaster* and *Penicillum chrysogenum* UGGT molecules was also a conclusion of the negative-stain EM studies in ref. 37.

Interdomain Flexibility Is Important for UGGT Activity. To investigate whether the observed interdomain flexibility of UGGT has a functional role, we engineered the double cysteine mutants *Ct*UGGT$^{D611C/G1050C}$ and *Ct*UGGT$^{N796C/G1118C}$, designed to form disulfide bridges across the TRXL2–βS2 and TRXL3–βS2 interfaces, respectively (Fig. 4*A*). The extra disulfide bridge in *Ct*UGGT$^{D611C/G1050C}$ disfavors both the open and intermediate conformations (Fig. 4*B*), whereas the one in *Ct*UGGT$^{N796C/G1118C}$ disfavors the open conformation (Fig. 4*C*). Mass spectrometry analysis of tryptic and peptic peptides of *Ct*UGGT$^{D611C/G1050C}$ and *Ct*UGGT$^{N796C/G1118C}$ confirms that the engineered disulfide bridges are indeed 100% formed (*SI Appendix*, Figs. S8 and S9). The *Ct*UGGT Cys double mutants show CD spectra equivalent to WT *Ct*UGGT, with the extra disulfide bridge increasing their T_m by 10 °C and 15 °C for *Ct*UGGT$^{N796C/G1118C}$ and *Ct*UGGT$^{D611C/G1050C}$, respectively (*SI Appendix*, Fig. S1 *F* and *G*). *Ct*UGGT$^{N796C/G1118C}$ and *Ct*UGGT$^{D611C/G1050C}$ display lower enzymatic activity than the wild-type enzyme in reglucosylating urea-denatured bovine thyroglobulin (Fig. 4*D*). The crystal structure of the *Ct*UGGT$^{D611C/G1050C}$ double mutant determined to 2.8-Å resolution (in magenta in Fig. 4*B*) shows clear electron density for the engineered disulfide bond. The GT domain in the *Ct*UGGT$^{D611C/G1050C}$ double mutant crystal structure shows no significant changes with respect to the WT structures (overall C_α rmsd = 1.04, 0.21, and 1.11 Å for 514 residues in the GT compared with the open, intermediate, and closed conformations, respectively), possibly because the intervening β-sandwiches βS1 and βS2 effectively shield the GT domain from the changes occurring at the distal end of the molecule. These observations rule out allostery propagating structural changes from the TRXL2–3 domains to the catalytic site. Taken together, our data suggest that efficient UGGT-mediated reglucosylation of misfolded glycoproteins depends on the observed UGGT interdomain flexibility. It is tempting to speculate that the movements of the TRXL domains of UGGT endow the enzyme with the promiscuity (52) needed to recognize and reglucosylate substrates of many sizes and shapes, and/or enables reglucosylation of N-linked glycans situated at variable distances from the site of misfold (33–35).

The structures of a eukaryotic UGGT presented in this work will inform further studies of its complexes with a number of misfolded client glycoproteins, to dissect the molecular determinants of UGGT substrate recognition and reglucosylation. The central role UGGT plays as the misfold sensor of the ER quality control (ERQC) machinery further makes it a potential pharmacological target in pathological conditions such as certain congenital protein folding diseases. This work paves the way to structure-based drug design of UGGT inhibitors that may have therapeutic potential for the rescue of the secretion of misfolded but functional glycoprotein mutants (20, 24).

Materials and Methods

Cloning. Amplification of the mature sequence of the *C. thermophilum* UGGT gene was accomplished by PCR and Gibson assembly (New England Biolabs) to insert into the expression vector pHLsec using standard protocols. Cloning of the double mutants *Ct*UGGT$^{N796C/G1118C}$ and *Ct*UGGT$^{D611C/G1050C}$: the coding sequence of *Ct*UGGT was cloned into the pDONR221 vector. The QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent) was used to introduce the mutations N796C/G1118C and D611C/G1050C, following manufacturer's instructions. The DNA double mutants were reinserted in pHLsec by Gibson assembly.

Protein Expression and Purification. Transfection into the FreeStyle 293 Expression System (Thermo Fisher Scientific) was carried out according to the manufacturer's protocol. After 5 d, the cells' supernatant was applied onto a Ni-affinity column equilibrated with PBS binding buffer. The protein was eluted with a linear gradient of imidazole. The concentrated enzyme was applied to a Superdex 200 16/600 column (GE Healthcare Life Sciences) in 20 mM Hepes pH 7.4, 140 mM NaCl.

Reglucosylation of Urea-Denatured Thyroglobulin. Bovine thyroglobulin (Sigma-Aldrich) was denatured with urea and reglucosylated with *Ct*UGGT following the protocol by Trombetta et al. (4). N-linked glycans were purified and detected as described in ref. 53. The amount of reglucosylation was measured by determining the peak area of the PNGase F released 2AA-labeled species $Man_9GlcNAc_2$ and $Glc_1Man_9GlcNAc_2$ using Waters Empower software.

Cryoelectron Microscopy. Cryomicroscopy specimens were prepared by applying 3 μL of *Ct*UGGT (0.25 mg/mL) to specimen grids and immediately plunged into liquid ethane. Electron micrographs were recorded on a Titan Krios microscope (FEI Company) operated at 300 kV using a Volta phase plate. Exposure movies were recorded using a Gatan K2 Summit detector. A total of 88,567 particles were down-sampled to 4 Å per pixel and subjected to three rounds of 2D classification in which bad classes were manually excluded from the dataset. A total of 33,819 images were used to generate an ab initio 3D starting model. One of the two hands of the ab initio model was used as a reference for 3D classification into four classes in RELION. Exclusion of TRXL2 from the mask used for alignment was necessary to see the resolution improvement from 21 to 15 Å, as judged by "gold-standard" Fourier shell correlation for the final 3D cryo-EM reconstruction consisting of 12,473 particles.

X-Ray Crystal Structures Determination. All sitting drops were set up at 21 °C. All crystals were harvested and flash frozen in liquid N_2. See *SI Appendix* for crystal growth conditions. X-ray diffraction data from the $P6_1$ and the *Ct*UGGT$^{D611C/G1050C}$ double mutant $P2_12_12_1$ crystals were collected on beamlines BM14 and ID30A-1, respectively, at the European Synchrotron Radiation Facility, Grenoble, France. X-ray diffraction data from the $P4_3$ and $P6_122$ crystal forms were collected on beamline I04-1 at the Diamond Light Source Harwell, UK. The crystal structure determination was as follows: for $P6_1$ crystal form, six Pt sites were found interpreting the anomalous difference Patterson maps from a K_2PtI_6 soaked crystal. Phased molecular replacement in the solvent flattened map allowed correct positioning of models for the TRXL2, TRXL3, and GT domains. The *Ct*UGGT $P6_122$, $P4_3$, and *Ct*UGGT$^{D611C/G1050C}$ double mutant $P2_12_12_1$ crystal forms were initially phased by molecular replacement in Phaser, searching with the $P6_1$ model without the TRXL2 domain. The *Ct*UGGT$^{D611C/G1050C}$ $P2_12_12_1$ model was refined with translation, libration, and screw (TLS) tensors restraints. The final $P6_1$, $P6_122$, and $P4_3$ model building and refinement was carried out using TLS and external restraints to the *Ct*UGGT$^{D611C/G1050C}$ $P2_12_12_1$ model.

Generation of Transgenic Plants. Binary vectors containing 35S::RFP–*Ct*UGGT and 35S::RFP were amplified in *Escherichia coli* and used for *Agrobacterium tumefaciens* (strain GV3101) transformation. Plant growth conditions were as follows: seedlings and plants of *A. thaliana* were grown at 22 °C and 70% relative humidity under a 16-h light/8-h dark cycle (light intensity ~120 μmol/m^{2}s).

The full, detailed methods for cloning, expression, purification, CD spectroscopy, activity assay, mass spectroscopy, cryo-EM, in planta confocal microscopy, SAXS, and X-ray crystal structure determination used in this study can be found in *SI Appendix, Materials and Methods*.

ACKNOWLEDGMENTS. We thank Chris Scanlan, Armando Parodi, Raymond Dwek, Maria Lucas, Eugene Valkov, Kathryn Scott, Patrizia Abrusci, Weston Struwe, and Mark Wormald for helpful discussions, advice, and comments on the manuscript; members of the N.Z. laboratory for assistance with molecular biology and protein chemistry; Radu Aricescu and Ed Hurt for donating the pHLsec vector and the *Ct*UGGT cDNA library, respectively; Ioannis Vakonakis for help with the DNA vector for expression in yeast;

David Staunton for assistance with the size-exclusion chromatography with multiple angle laser light scattering and CD measurements; Elena Seiradake for donating GST-Endo F1 glycosidase; Ed Lowe and the staff at beamline I04-1 at the Diamond Light Source (DLS) (Harwell, UK) and beamlines BM14-U, BM29, and ID30A-1 at the European Synchrotron Radiation Facility (Grenoble, France) for assistance with X-ray and SAXS data collection; Oliver Clarke, Luigi De Colibus, Clemens Vonrhein, and Claus Flensburg for advice about model building and refinement; Alistair Siebert (DLS) and Felix de Haas (FEI Company) for advice and assistance with electron microscopy data collection; Colin Palmer, Tom Burnley (Collaborative Computational Project in Electron Microscopy, Research Complex at Harwell), and Tina Fredrick for computing support at DLS; Jianming Li for donating the seeds of *A. thaliana ebs1-3/bri1-9* mutant plants; Miriam Aber Schimera for help with the cloning of *Ct*UGGT into vectors for plant transfection; Svenja Hester for help with mass spectrometry of the double Cys mutants; and DLS for access and support to Electron Bio-Imaging Centre. A.T.C. and J.C.H. were funded by Wellcome Trust 4-Year Studentships 097300/Z/11/Z and 106272/Z/14/Z, respectively. N.Z. is a Fellow of Merton College, Oxford.

1. Apweiler R, Hermjakob H, 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. Hammond C, Braakman I, 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.
3. Parodi DAJ, Caramelo JJ, D'Alessio C (2014) *UDP-Glucose: Glycoprotein Glucosyltransferase 1,2 (UGGT1,2). Handbook of Glycosyltransferases and Related Genes* (Springer Japan, Tokyo), pp 15–30.
4. Trombetta SE, Bosch M, Parodi AJ (1989) Glucosylation of glycoproteins by mammalian, plant, fungal, and trypanosomatid protozoa microsomal membranes. *Biochemistry* 28:8108–8116.
5. Sousa MC, Ferrero-Garcia MA, Parodi AJ (1992) Recognition of the oligosaccharide and protein moieties of glycoproteins by the UDP-Glc:glycoprotein glucosyltransferase. *Biochemistry* 31:97–105.
6. Gardner TG, Kearse KP (1999) Modification of the T cell antigen receptor (TCR) complex by UDP-glucose:glycoprotein glucosyltransferase. TCR folding is finalized convergent with formation of alpha beta delta epsilon gamma epsilon complexes. *J Biol Chem* 274: 14094–14099.
7. Parodi AJ, Mendelzon DH, Lederkremer GZ (1983) Transient glucosylation of protein-bound Man9GlcNAc2, Man8GlcNAc2, and Man7GlcNAc2 in calf thyroid cells. A possible recognition signal in the processing of glycoproteins. *J Biol Chem* 258:8260–8265.
8. Parodi AJ, Mendelzon DH, Lederkremer GZ, Martin-Barrientos J (1984) Evidence that transient glucosylation of protein-linked Man9GlcNAc2, Man8GlcNAc2, and Man7GlcNAc2 occurs in rat liver and Phaseolus vulgaris cells. *J Biol Chem* 259:6351–6357.
9. Michalak M, Corbett EF, Mesaeli N, Nakamura K, Opas M (1999) Calreticulin: one protein, one gene, many functions. *Biochem J* 344(Pt 2):281–292.
10. McCaffrey K, Braakman I (2016) Protein quality control at the endoplasmic reticulum. *Essays Biochem* 60:227–235.
11. Molinari M, Galli C, Vanoni O, Arnold SM, Kaufman RJ (2005) Persistent glycoprotein misfolding activates the glucosidase II/UGT1-driven calnexin cycle to delay aggregation and loss of folding competence. *Mol Cell* 20:503–512.
12. Huang P-N, et al. (2017) UGGT1 enhances enterovirus 71 pathogenicity by promoting viral RNA synthesis and viral replication. *PLoS Pathog* 13:e1006375.
13. Hebert DN, Lamriben L, Powers ET, Kelly JW (2014) The intrinsic and extrinsic effects of N-linked glycans on glycoproteostasis. *Nat Chem Biol* 10:902–910.
14. Soldà T, Galli C, Kaufman RJ, Molinari M (2007) Substrate-specific requirements for UGT1-dependent release from calnexin. *Mol Cell* 27:238–249.
15. Lamriben L, Graham JB, Adams BM, Hebert DN (2016) N-glycan-based ER molecular chaperone and protein quality control system: The calnexin binding cycle. *Traffic* 17:308–326.
16. Blanco-Herrera F, et al. (2015) The UDP-glucose: Glycoprotein glucosyltransferase (UGGT), a key enzyme in ER quality control, plays a significant role in plant growth as well as biotic and abiotic stress in Arabidopsis thaliana. *BMC Plant Biol* 15:127.
17. Neerincx A, et al. (2017) TAPBPR bridges UDP-glucose:glycoprotein glucosyltransferase 1 onto MHC class I to provide quality control in the antigen presentation pathway. *eLife* 6:e23049.
18. Zhang W, Wearsch PA, Zhu Y, Leonhardt RM, Cresswell P (2011) A role for UDP-glucose glycoprotein glucosyltransferase in expression and quality control of MHC class I molecules. *Proc Natl Acad Sci USA* 108:4956–4961.
19. Parenti G, Andria G, Valenzano KJ (2015) Pharmacological chaperone therapy: Preclinical development, clinical translation, and prospects for the treatment of lysosomal storage disorders. *Mol Ther* 23:1138–1148.
20. Merulla J, Soldà T, Molinari M (2015) A novel UGGT1 and p97-dependent checkpoint for native ectodomains with ionizable intramembrane residue. *Mol Biol Cell* 26:1532–1542.
21. Tannous A, Patel N, Tamura T, Hebert DN (2015) Reglucosylation by UDP-glucose: glycoprotein glucosyltransferase 1 delays glycoprotein secretion but not degradation. *Mol Biol Cell* 26:390–405.
22. Pankow S, et al. (2015) ΔF508 CFTR interactome remodelling promotes rescue of cystic fibrosis. *Nature* 528:510–516.
23. Trombetta ES, Helenius A (1999) Glycoprotein reglucosylation and nucleotide sugar utilization in the secretory pathway: Identification of a nucleoside diphosphatase in the endoplasmic reticulum. *EMBO J* 18:3282–3292.
24. Amara JF, Cheng SH, Smith AE (1992) Intracellular protein trafficking defects in human disease. *Trends Cell Biol* 2:145–149.
25. Guerin M, Parodi AJ (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.
26. Arnold SM, Kaufman RJ (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.
27. Albesa-Jové D, Guerin ME (2016) The conformational plasticity of glycosyltransferases. *Curr Opin Struct Biol* 40:23–32.
28. Zhu T, Satoh T, 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.
29. Labriola C, Cazzulo JJ, Parodi AJ (1999) Trypanosoma cruzi calreticulin is a lectin that binds monoglucosylated oligosaccharides but not protein moieties of glycoproteins. *Mol Biol Cell* 10:1381–1394.
30. Jin H, Yan Z, Nam KH, Li J (2007) Allele-specific suppression of a defective brassinosteroid receptor reveals a physiological role of UGGT in ER quality control. *Mol Cell* 26:821–830.
31. Li J, et al. (2009) Specific ER quality control components required for biogenesis of the plant innate immune receptor EFR. *Proc Natl Acad Sci USA* 106:15973–15978.
32. Pearse BR, et al. (2010) The role of UDP-Glc:glycoprotein glucosyltransferase 1 in the maturation of an obligate substrate prosaposin. *J Cell Biol* 189:829–841.
33. Ritter C, Helenius A (2000) Recognition of local glycoprotein misfolding by the ER folding sensor UDP-glucose:glycoprotein glucosyltransferase. *Nat Struct Biol* 7:278–280.
34. Taylor SC, Ferguson AD, Bergeron JJM, Thomas DY (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.
35. Ritter C, Quirin K, Kowarik M, Helenius A (2005) Minor folding defects trigger local modification of glycoproteins by the ER folding sensor GT. *EMBO J* 24:1730–1738.
36. Amlacher S, et al. (2011) Insight into structure and assembly of the nuclear pore complex by utilizing the genome of a eukaryotic thermophile. *Cell* 146:277–289.
37. Calles-Garcia D, et al. (2017) Single-particle electron microscopy structure of UDP-glucose:glycoprotein glucosyltransferase suggests a selectivity mechanism for misfolded proteins. *J Biol Chem*:jbc.M117.789495.
38. Arnold SM, Fessler LI, Fessler JH, Kaufman RJ (2000) Two homologues encoding human UDP-glucose:glycoprotein glucosyltransferase differ in mRNA expression and enzymatic activity. *Biochemistry* 39:2149–2163.
39. Sevvana M, et al. (2006) Structural elucidation of the PDI-related chaperone Wind with the help of mutants. *Acta Crystallogr D Biol Crystallogr* 62:589–594.
40. Lippert U, Diao D, Barak NN, Ferrari DM (2007) Conserved structural and functional properties of D-domain containing redox-active and -inactive protein disulfide isomerase-related protein chaperones. *J Biol Chem* 282:11213–11220.
41. Barak NN, et al. (2009) Crystal structure and functional analysis of the protein disulfide isomerase-related protein ERp29. *J Mol Biol* 385:1630–1642.
42. Funkner A, et al. (2013) Peptide binding by catalytic domains of the protein disulfide isomerase-related protein ERp46. *J Mol Biol* 425:1340–1362.
43. Kozlov G, Määttänen P, Thomas DY, Gehring K (2010) A structural overview of the PDI family of proteins. *FEBS J* 277:3924–3936.
44. Vinaik R, Kozlov G, Gehring K (2013) Structure of the non-catalytic domain of the protein disulfide isomerase-related protein (PDIR) reveals function in protein binding. *PLoS One* 8:e62021.
45. Takeda Y, et al. (2016) Effects of domain composition on catalytic activity of human UDP-glucose:glycoprotein glucosyltransferases. *Glycobiology* 26:999–1006.
46. Banerjee S, et al. (2007) The evolution of N-glycan-dependent endoplasmic reticulum quality control factors for glycoprotein folding and degradation. *Proc Natl Acad Sci USA* 104:11676–11681.
47. Samuelson J, Robbins PW (2015) Effects of N-glycan precursor length diversity on quality control of protein folding and on protein glycosylation. *Semin Cell Dev Biol* 41:121–128.
48. Edelhoch H, Lippoldt RE (1962) The properties of thyroglobulin. IX. The molecular properties of iodinated thyroglobulin. *J Biol Chem* 237:2788–2794.
49. Zuber C, et al. (2001) Immunolocalization of UDP-glucose:glycoprotein glucosyltransferase indicates involvement of pre-Golgi intermediates in protein quality control. *Proc Natl Acad Sci USA* 98:10710–10715.
50. Vert G, Nemhauser JL, Geldner N, Hong F, Chory J (2005) Molecular mechanisms of steroid hormone signaling in plants. *Annu Rev Cell Dev Biol* 21:177–201.
51. Parker CG, Fessler LI, Nelson RE, Fessler JH (1995) Drosophila UDP-glucose:glycoprotein glucosyltransferase: sequence and characterization of an enzyme that distinguishes between denatured and native proteins. *EMBO J* 14:1294–1303.
52. Nobeli I, Favia AD, Thornton JM (2009) Protein promiscuity and its implications for biotechnology. *Nat Biotechnol* 27:157–167.
53. Caputo AT, et al. (2016) Structures of mammalian ER α-glucosidase II capture the binding modes of broad-spectrum iminosugar antivirals. *Proc Natl Acad Sci USA* 113: E4630–E4638.