JBC RESEARCH ARTICLE

Structural insights into how GlcNAc-1-phosphotransferase directs lysosomal protein transport

Received for publication, August 19, 2021, and in revised form, February 3, 2022 Published, Papers in Press, February 9, 2022,
https://doi.org/10.1016/j.jbc.2022.101702

Shuo Du (杜硕),[1,2,‡], Guopeng Wang (王国鹏),[2,‡], Zhiying Zhang (张志莹),[1,2], Chengying Ma (马成英),[2,3,4], Ning Gao (高宁),[2,3,4,*], and Junyu Xiao (肖俊宇)[1,2,4,5,*]

From the [1]State Key Laboratory of Protein and Plant Gene Research, [2]School of Life Sciences, [3]State Key Laboratory of Membrane Biology, [4]Peking-Tsinghua Center for Life Sciences, and [5]Beijing Advanced Innovation Center for Genomics, Peking University, Beijing, China

Edited by Gerald Hart

GlcNAc-1-phosphotransferase catalyzes the initial step in the formation of the mannose-6-phosphate tag that labels ~60 lysosomal proteins for transport. Mutations in GlcNAc-1-phosphotransferase are known to cause lysosomal storage disorders such as mucolipidoses. However, the molecular mechanism of GlcNAc-1-phosphotransferase activity remains unclear. Mammalian GlcNAc-1-phosphotransferases are α2β2γ2 hexamers in which the core catalytic α- and β-subunits are derived from the *GNPTAB* (N-acetylglucosamine-1-phosphate transferase subunits alpha and beta) gene. Here, we present the cryo-electron microscopy structure of the *Drosophila melanogaster* GNPTAB homolog, DmGNPTAB. We identified four conserved regions located far apart in the sequence that fold into the catalytic domain, which exhibits structural similarity to that of the UDP–glucose glycoprotein glucosyltransferase. Comparison with UDP–glucose glycoprotein glucosyltransferase also revealed a putative donor substrate-binding site, and the functional requirements of critical residues in human GNPTAB were validated using *GNPTAB*-knockout cells. Finally, we show that DmGNPTAB forms a homodimer that is evolutionarily conserved and that perturbing the dimer interface undermines the maturation and activity of human GNPTAB. These results provide important insights into GlcNAc-1-phosphotransferase function and related diseases.

Protein phosphorylation is universally present as a regulatory strategy in eukaryotic cells. Glycan phosphorylation, though not as abundant, also plays essential roles in modulating cellular processes, particularly within the secretory compartments. A canonical glycan phosphorylation event involves ~60 secretory proteins that are destined for lysosomes. Similar to other secretory molecules, these lysosomal proteins are first synthesized in the endoplasmic reticulum and then traverse through the Golgi network. At the Golgi apparatus, these proteins are "phosphorylated" on a terminal mannose residue in their N-linked glycans, resulting in the formation of a mannose 6-phosphate (M6P) tag that is recognized by two specific M6P receptors to direct their lysosomal transport. Interestingly, this essential modification is not performed by an ATP-dependent kinase but is generated by the sequential action of two enzymes: first, the N-acetylglucosamine-1-phosphotransferase (GlcNAc-1-phosphotransferase) catalyzes the addition of an N-acetylglucosamine-1-phosphate group to the terminal mannose, and the GlcNAc-1-phosphodiester α-N-acetylglucosaminidase then removes the GlcNAc moiety to uncover M6P (1).

Mammalian GlcNAc-1-phosphotransferases are large protein complexes that comprise two α-subunits, two β-subunits, and two γ-subunits (2, 3). The catalytic α- and β-subunits are first synthesized as GNPTAB fusion proteins. GNPTAB has a complex structural organization and includes two transmembrane segments, four conserved regions (CR1–CR4), two Notch repeats (N1 and N2), a DNA methyltransferase-associated protein (DMAP) interaction domain, and several spacer regions (S1–S4) (Fig. 1*A*). The two transmembrane segments anchor GNPTAB on the Golgi membrane and project most of the molecule into the Golgi lumen. The four CRs are conserved in Stealth proteins, and members of this family can function as hexose-phosphate transferases in bacteria to synthesize cell wall polysaccharides (4). The Notch repeats and DMAP interaction domain mediate the interaction between GlcNAc-1-phosphotransferase and its diverse lysosomal protein substrates (5–7). The spacer regions are also functionally important. For example, the S2 region is responsible for interacting with the γ-subunit (8, 9). S3 contains a recognition site for the site-1 protease, which cleaves GNPTAB into its α- and β-subunits and thereby leads to catalytic activation (10). The S1 spacer facilitates proper processing of GNPTAB by the site-1 protease (11). S4 contains an EF-hand calcium-binding motif, but its function remains unclear. The γ-subunit is encoded by *GNPTG* and plays a regulatory function to promote the activity of the GlcNAc-1-phosphotransferase holoenzyme toward a subset of substrates (12).

More than 200 mutations in GlcNAc-1-phosphotransferase have been documented in patients with mucolipidoses, a group

‡ These authors contributed equally to this work.
* For correspondence: Junyu Xiao, junyuxiao@pku.edu.cn; Ning Gao, gaon@pku.edu.cn.

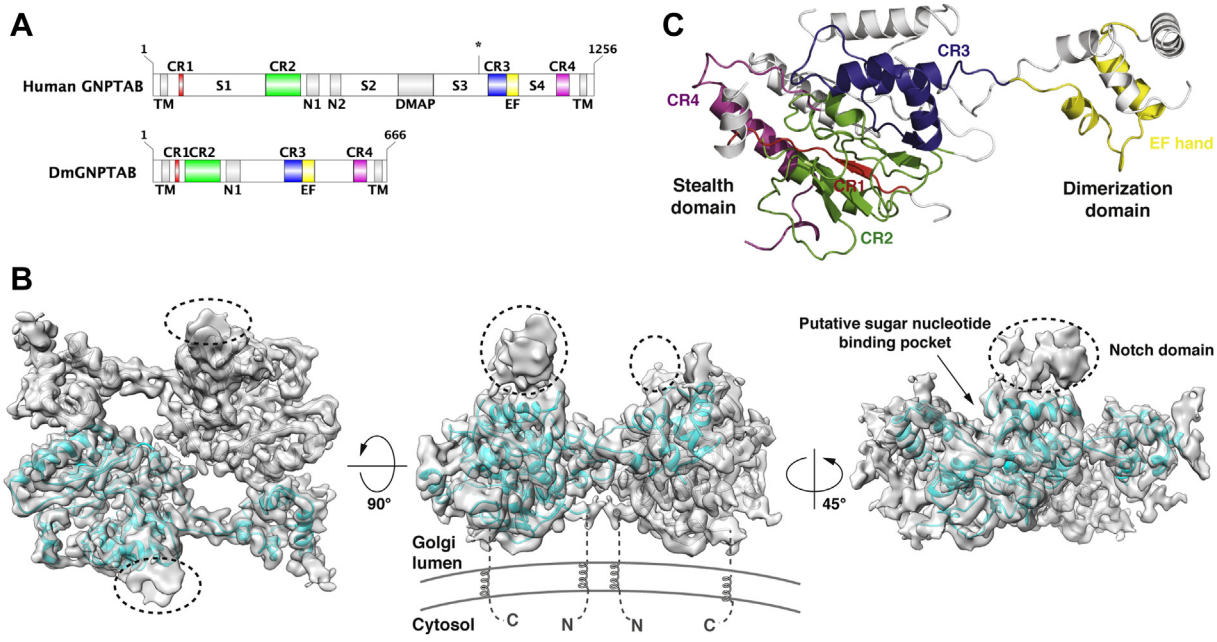

Figure 1. Cryo-EM structure of DmGNPTAB. *A*, domain architectures of human GNPTAB and DmGNPTAB. The *asterisk* indicates the S1P cleavage site in human GNPTAB. *B*, cryo-EM density map of DmGNPTAB. The structure model is shown inside the map, with the two protomers displayed in *cyan* and *gray*, respectively. The Notch domain regions (*dashed circles*) displayed weak densities and could not be modeled. Putative model of DmGNPTAB on the Golgi membrane is illustrated in the *middle panel*, with the N and C termini of each molecule indicated. *C*, ribbon diagram of the DmGNPTAB monomer. The four CRs and EF hand motif in one protomer are depicted in *red*, *green*, *blue*, *magenta*, and *yellow*, respectively, whereas, the rest of this protomer is shown in *white*. CR1 to CR4, conserved regions in the Stealth proteins; cryo-EM, cryo-electron microscopy; DMAP, DNA methyltransferase-associated protein interaction domain; N1 and N2, Notch repeats; S1 to S4, spacer regions; TM, transmembrane segment.

of human lysosomal storage disorders that include skeletal and neuronal abnormalities (summarized in (13)). Mutations in GlcNAc-1-phosphotransferase have also been linked to stuttering (14). On the other hand, GlcNAc-1-phosphotransferase might serve as a potential antiviral target because a number of viruses, such as the Ebola virus and the common cold coronaviruses OC43 and 229E, rely on the lysosomal pathway for infection and egress (15, 16). However, the underlying molecular mechanism of GlcNAc-1-phosphotransferase remains insufficiently understood. Here, we sought to characterize the structural architecture of GlcNAc-1-phosphotransferase by cryo-electron microscopy (cryo-EM) and successfully determined the 3.5-Å structure of the *Drosophila melanogaster* GNPTAB. Our results reveal critical structural features that are conserved in the GNPTAB family. We then generated a *GNPTAB*-knockout cell line using the CRISPR-Cas9 genome editing technique and validated the importance of residues in human GNPTAB involved in donor substrate binding and dimerization. We also analyzed pathogenic missense mutations and assessed their potential impacts. Together, our results advance the understanding of GlcNAc-1-phosphotransferase and related human diseases.

Results and discussion

Cryo-EM structure of D. melanogaster GNPTAB

We sought to investigate the structural basis of GlcNAc-1-phosphotransferase. Despite intensive attempts, we were unable to obtain the structure of a mammalian GlcNAc-1-phosphotransferase and could not determine the structure of the α/β subcomplex or the GNPTAB precursor. *D. melanogaster* has a GNPTAB homolog (DmGNPTAB) but lacks a discernable gene encoding GNPTG (4, 17). The DmGNPTAB protein is markedly more compact than its human counterpart (Fig. 1*A*). Specifically, this protein contains all four CRs and a similar CR3-CR4 spacer but has shorter CR1-CR2 and CR2-CR3 spacers, only one Notch repeat, and lacks the DMAP interaction domain and S1P cleavage site (Figs. 1*A* and S1). We obtained the luminal portion of DmGNPTAB and analyzed its structure by cryo-EM (Figs. 1*B* and S2). The structure was determined at an overall resolution of 3.5 Å (Table 1). The center region displayed high resolutions, which allowed us to build the structural model *de novo*. Approximately half of the protein molecule, including all four CRs and the CR3-CR4 spacer, could be confidently placed. The rest of the molecule, particularly the CR2-CR3 spacer including the Notch module, displayed weak densities and was thus not modeled. The amino and carboxyl termini are located on the same side of a DmGNPTAB monomer, which sheds light on the topology of the full-length protein dimer on the Golgi membrane (Fig. 1*B*). The membrane localization of mammalian GNPTAB is likely important for its proper processing by membrane-bound S1P within the Golgi apparatus (10).

Two domains can be clearly discerned in the structural model of the DmGNPTAB monomer (Fig. 1*C*). The large domain, which is referred to as the Stealth domain, features an α/β fold that includes all four Stealth CRs. CR1 contains a β-strand that occupies the center of the Stealth domain. CR2

Table 1
Cryo-EM data collection, processing, and validation statistics

	GNPTAB (EMDB & PDB IDs:)
Data collection and processing	
Voltage (kV)	300
Microscope	FEI Titan Krios G3
Camera	K2 Summit (Gatan)
Magnification (calibrated)	210,000×
Electron exposure ($e^-/Å^2$)	60.29
Exposure rate ($e^-/Å^2/s$)	18.84
Number of frames collected per micrograph	32
Energy filter slit width	20 eV
Automation software	SerialEM
Defocus range (μm)	−1.0 to −2.2
Pixel size (Å)	0.6516
Micrographs used	14,345
Estimated accuracy of rotations	2.447
Symmetry imposed	C2
Initial particle images	5,768,795
Final particle images	131,301
Resolution at 0.143 FSC of masked reconstruction (Å)	3.53
Refinement	
Initial model used (PDB code)	none
Refinement package	Phenix v1.18.2 (Real-space refinement at 3.53 Å)
Map-model CC	
CC_mask	0.81
CC_box	0.73
CC_peaks	0.64
CC_volume	0.79
Model composition	
Nonhydrogen atoms	6,130
Protein residues	742
Ligands	2
B factors ($Å^2$)	
Protein	24.82
Ligands	62.76
R.m.s. deviations	
Bond lengths (Å)	0.006
Bond angles (°)	1.292
Validation	
MolProbity score	1.89
Clashscore	4.71
Poor rotamers (%)	0.30
Ramachandran plot	
Favored (%)	85.54
Allowed (%)	13.91
Disallowed (%)	0.55
Cβ outliers (%)	0.00

Abbreviations: cryo-EM, cryo-electron microscopy; FSC, Fourier shell correlation.

comprises three β-strands and two α-helices. The three strands sandwich the CR1 strand in a parallel manner to form a β-sheet, whereas the two helices are situated on each side of the sheet. CR3 consists of one strand and three helices. The single CR3 strand runs antiparallel to the four strands described above, and the three helices bundle together with the short helix in CR2. CR4 features a single helix, which packs against the long helix in CR2. The small domain is encoded by the CR3-CR4 spacer, including the calcium-binding EF hand motif, and features a helix bundle to mediate dimerization.

Structure of the stealth domain

The four Stealth CRs, which are located far apart in the sequence, fold into a single globular domain. Structural homology search suggests that the Stealth domain displays structural similarities to several GT-A-type glycosyltransferases (18) despite low sequence homology. In particular, the Stealth domain exhibits structural resemblance to UDP–glucose glycoprotein glucosyltransferase (UGGT), particularly at the central β-sheet region (Fig. 2, *A* and *B*). UGGT is an ER-resident protein that surveils the folding status of secretory glycoproteins (19). Similar to GlcNAc-1-phosphotransferase, UGGT is a label marker, and the labeling of UGGT also occurs on the terminal mannose of an N-linked glycan. Instead of labeling proteins that are destined for lysosomes, UGGT labels proteins that are incompletely folded; specifically, it recognizes misfolded glycoproteins and transfers a glucose residue to the terminal mannose on the glycan of these proteins, and this modification is recognized by ER chaperones, including calnexin and calreticulin, to facilitate correct folding.

Crystal structures of UGGT homologs from several thermophilic fungi have been determined (20, 21). A structural comparison between the Stealth domain of DmGNPTAB and the catalytic domain of *Thermomyces dupontii* UGGT (TdUGGT) in complex with UDP-glucose offers a glimpse into the sugar nucleotide-binding site of DmGNPTAB. UDP-glucose binds to the surface pocket of TdUGGT (Fig. 2*A*). DmGNPTAB has a similar surface pocket, lined by a group of conserved residues, including Thr69 from CR1; Ser156, Ile159, Glu160, Tyr175, Asn177, Asp178, and Asp179 from CR2; His375, Phe378, Arg405, and Gln411 from CR3; Phe546 and Met548 from the CR3-CR4 spacer; and Cys572, Asn574, and Asn576 from CR4 (Fig. 2, *B* and *C*). The functional importance of these residues is underscored by the fact that missense mutations of a number of the corresponding human residues have been found in patients (see below). A Ca^{2+} ion facilitates the accommodation of UDP-glucose in TdUGGT and is coordinated by three Asp residues, including Asp1294 and Asp1296 in the Asp-X-Asp signature motif of GT-A glycosyltransferases and Asp1427 (Fig. 2*D*). A well-conserved Asn177-Asp178-Asp179 motif in CR2 of DmGNPTAB (Fig. S1) aligns with the Asp-X-Asp motif in TdUGGT, whereas Cys572 appears to take the position of TdUGGT-Asp1427. It is thus likely that these residues are also involved in the binding to a divalent cation that assists in positioning the UDP-GlcNAc in DmGNPTAB.

The Notch repeat domain displays weak density and cannot be unambiguously modeled. Nevertheless, it appears to be located in the vicinity of the sugar nucleotide-binding pocket and opposite the side that faces the membrane (Fig. 1*B*). The Notch domains of human GNPTAB are involved in binding to the lysosomal protein substrates (6, 7). Based on the structural comparison with TdUGGT (Fig. 2*B*), we envisioned that the donor substrate UDP-GlcNAc is likely accommodated in DmGNPTAB in an orientation similar to that of UDP-glucose in TdUGGT, with the N-acetylglucosamine group holding toward the direction of the Notch domain that engages the cognate acceptor substrates in *Drosophila*.

Human GNPTAB mutants are functionally defective

We sought to validate the functional importance of some of these residues in human GNPTAB. To unambiguously analyze the activities of various GNPTAB mutants, we first generated a *GNPTAB*-knockout ($GNPTAB^{-/-}$) HeLa cell line using the CRISPR/Cas9 genome-editing technique (Fig. 3*A*). Consistent with previous observations (6), $GNPTAB^{-/-}$ cells exhibited

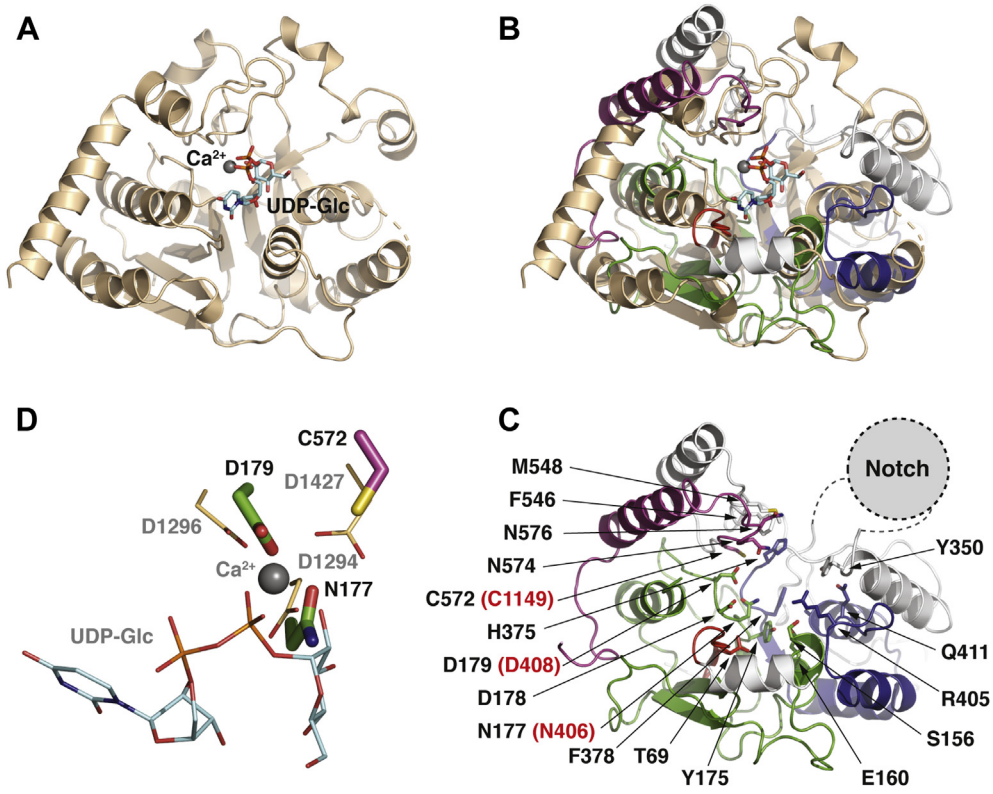

Figure 2. Structure of the Stealth domain. *A*, crystal structure of TdUGGT in complex with Ca^{2+}/UDP-glucose (PDB ID: 5H18). *B*, structural superposition of DmGNPTAB and TdUGGT shown in the same orientation as in panel *A*. DmGNPTAB is shown using the same color scheme as in Figure 1*C*. *C*, putative UDP-GlcNAc-binding pocket. The Stealth domain is shown in the same orientation as in panel *A* and *B*. Residues involved in forming the putative UDP-GlcNAc-binding pocket are highlighted. Human GNPTAB residues N406, D408, and C1149 correspond to N177, D179, and C572 in DmGNPTAB. *D*, Asn177, Asp179, and Cys572 in DmGNPTAB appear to align with Asp1294, Asp1296, and Asp1427 in TdUGGT, which coordinate a Ca^{2+} ion to position UDP-Glc in TdUGGT. TdUGGT, *Thermomyces dupontii* UGGT; UGGT, UDP–glucose glycoprotein glucosyltransferase.

markedly swollen lysosomes compared with the parental cells (Fig. 3*B*), which indicated that lysosomal function is severely impaired in these cells. As a result of lysosomal dysfunction, the lysosomal cysteine protease cathepsin B (CatB) was not properly processed, and the mature form of endogenous CatB at ∼30 kDa was not detected in the lysates of *GNPTAB*$^{-/-}$ cells (Fig. 3*C*). This finding was also consistent with previous observations obtained with GNPTAB-deficient HAP1 cells (15). We then generated alanine substitutions of Asn406, Asp408, and Cys1149, which are equivalent to Asn177, Asp179, and Cys572 in DmGNPTAB that form the putative metal-binding site (Fig. 2*D*) and examined the abilities of these mutants to rescue CatB maturation. As expected, the expression of WT GNPTAB restored the mature form of CatB (Fig. 3*D*). In contrast, N406A, D408A, and C1149A displayed reduced activities compared with the WT protein, and D408A appeared to be completely inactive. These results demonstrate the importance of these residues for the function of GlcNAc-1-phosphotransferase, corroborating our structural analyses.

A conserved dimeric architecture

Mammalian GlcNAc-1-phosphotransferases are α2β2γ2 hexameric complexes, and the construction of the hexamer remains poorly understood. In our structure, two DmGNPTAB molecules form a homodimer that resembles two fishes nestling against each other in a head-to-tail orientation (Fig. 1*B*). The dimer is mainly mediated by the CR3-CR4 spacer and CR4, whose counterparts both reside in the β-subunit of human GlcNAc-1-phosphotransferase (Fig. 1*A*). A few residues in CR2 also contribute to dimer formation. The dimer interface buries ∼1500-Å^2 solvent-accessible surfaces from each molecule and involves a number of invariant residues (Fig. S1), which suggests that the dimerization mechanism observed in this study is likely generally conserved in the GNPTAB family.

To verify the functional relevance of the dimer, we generated two human GNPTAB mutants: G1 (T1019N/D1020G/Q1021S) and G2 (T1035N/R1036G/I1037S). These two mutants were designed to create sites that allow the attachment of N-linked glycans. Thr1019 and Thr1035 in human GNPTAB correspond to Thr438 and Thr454 in DmGNPTAB, both of which are located in the dimer interface (Fig. 4*A*). The introduction of bulky glycans at these positions would impede dimerization of the human protein. We tagged these mutants with Flag tags at the C terminus, co-expressed them with V5-tagged WT GNPTAB in HEK293T cells, and performed Flag immunoprecipitation. Flag-tagged and V5-tagged WT GNPTAB proteins were efficiently coprecipitated (Fig. 4*B*), which indicated the formation of GNPTAB dimers in the cells. In contrast, the interactions between G1 or G2 and WT GNPTAB were markedly reduced, which suggested that these two mutants exhibit decreased abilities to dimerize with the WT protein. Furthermore, unlike WT GNPTAB, these two

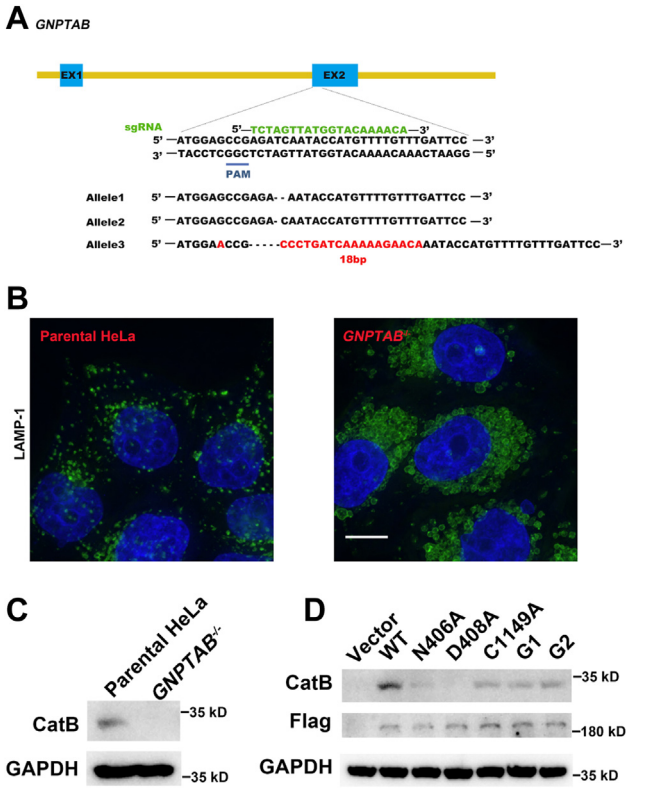

Figure 3. Human GNPTAB mutants are functionally defective. *A*, generation of the *GNPTAB*-knockout HeLa cell line by CRISPR-Cas9 genome editing. The sequences of the mutated alleles are shown. *B*, confocal immunofluorescence images of parental and GNPTAB-deficient HeLa cells stained using the late endosomal/lysosomal marker LAMP-1 (*green*). Nuclei are stained with DAPI (*blue*). The scale bar is 10 μm. *C*, mature cathepsin B (CatB) was not detected in GNPTAB-deficient cells. *D*, GNPTAB mutants could not fully rescue CatB processing in GNPTAB-deficient cells. The expression of GNPTAB was monitored using anti-Flag antibody. G1 and G2 stand for T1019N/D1020G/Q1021S and T1035N/R1036G/I1037S, respectively.

mutants were not efficiently processed because the ∼48-kDa band that corresponds to the β-subunit was not observed after their expression (Fig. 4*B*). Importantly, these mutations could also not fully rescue the maturation of CatB in *GNPTAB*$^{-/-}$ cells (Fig. 3*D*). Together, these results demonstrated that proper dimer formation is important for the processing of GNPTAB and for the activity of the GlcNAc-1-phosphotransferase holoenzyme.

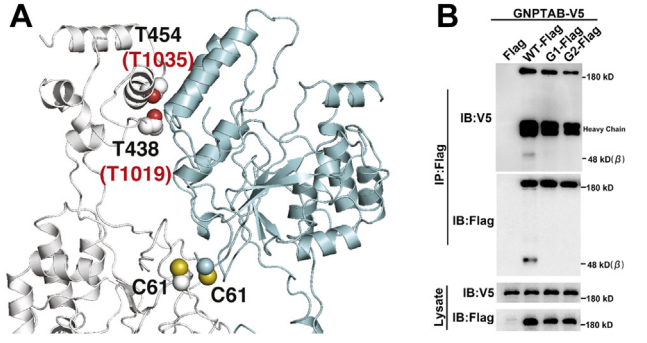

Figure 4. Mutations in the dimer interface perturb human GNPTAB dimerization and maturation. *A*, structure of the DmGNPTAB dimer. Cys61, Thr438, and Thr454 are highlighted. *B*, G1 (T1019N/D1020G/Q1021S) and G2 (T1035N/R1036G/I1037S) displayed defective processing and diminished abilities to form heterodimers with WT GNPTAB.

The CR3-CR4 spacer and CR4 play dominant roles in mediating the formation of the DmGNPTAB homodimer, which likely reflects how the β2 dimer is formed in human GlcNAc-1-phosphotransferase. In addition to the interactions between β-subunits, Cys70 in human GNPTAB is involved in disulfide-linked homodimerization of the α-subunits (8). Cys61 in DmGNPTAB appears to align with Cys70, and in the DmGNPTAB dimer, two Cys61 residues from the two monomers are located in close proximity (Fig. 4*A*), which suggests that a disulfide bond between the two corresponding Cys70 in human GNPTAB could be readily formed to further stabilize the α2β2 subcomplex. The γ-subunit also forms a disulfide-linked homodimer (3), and only a γ2 dimer can be assembled into the GlcNAc-1-phosphotransferase holoenzyme (22). The γ2 dimer, through its interactions with the S2 spacers in the two α-subunits, supplies another layer of interaction to architect the α2β2γ2 hexamer.

Insights into human mucolipidoses

Our results provide molecular insights into the pathogenesis of human mucolipidoses caused by GNPTAB mutations. More than 200 mutations in GNPTAB, including at least 65 missense mutations, have been detected in patients with various forms of mucolipidoses (13). The effects of some of these mutations have been characterized biochemically, but the structure of human GNPTAB has remained elusive despite its long research history. The structure of DmGNPTAB reveals a dimeric framework that sheds light on the structural core of human GlcNAc-1-phosphotransferase and can be used to further assess the molecular impacts of disease-causing mutations (Fig. 5). For example, the human GNPTAB residues Ser385, Glu389, Asp407, Asp408, His956, Arg986, and Asn1153 correspond to Ser156, Glu160, Asp178, Asp179, His375, Arg405, and Asn576 in DmGNPTAB (Fig. 5*A*), all of which are found among the sugar nucleotide-binding pockets described above (Fig. 2*C*). Thus, their mutations likely affected the binding of UDP-GlcNAc, consistent with the findings from previous biochemical studies showing that missense mutations involving these sites, such as S385L, E389K, D407A, D408N, H956Y, R986C, and N1153, resulted in markedly decreased enzyme activities (7, 23, 24). We also showed that D408A was unable to restore the maturation of CatB in *GNPTAB*$^{-/-}$ cells (Fig. 3*D*). Another set of mutants, including W81L, R334Q and R334L, S399F, I403T, and D1018G, displayed defects in exit from the ER, which indicated protein misfolding (7). These sites were also conserved in DmGNPTAB (Trp70, Arg105, Ser170, Leu174, and Asp437; Fig. 5*A*). Trp70 and Leu174 are both present in the interior of the structure and are intimately packed with surrounding hydrophobic residues. Arg105 forms a bidentate interaction with Glu607 to support Pro601, which is involved in dimerization. Ser170 likely forms hydrogen bond interactions with Asp65 and Arg115, and the corresponding human residues, Asp76 and Arg344, are also mutated in some patients (Fig. 5*A*). Similarly, some of the other mutations also lead to structural disturbance of human GlcNAc-1-phosphotransferase, as can be rationalized by the DmGNPTAB structure.

A

Select human GNPTAB missense mutations	Corresponding residues in DmGNPT
Asp76Gly	Asp65
Trp81Leu	Trp70
Tyr331Cys	Tyr102
Arg334Leu, Arg334Gln, Arg334Pro	Arg105
Arg344Trp	Arg115
Ile348Leu	Ile119
Thr382Pro	Thr153
Ser385Leu	Ser156
Glu389Lys	Glu160
His393Arg	His164
Ser399Phe	Ser170
Ile403Thr	Leu174
Asp407Ala	Asp178
Asp408Asn	Asp179
Ala955Val	Ala374
His956Tyr, His956Arg	His375
Arg986Cys, Arg986Gly, Arg986His	Arg405
Tyr997Cys	Tyr416
Leu1001Pro	Leu420
Asp1018Gly, Asp1018Val	Asp437
Asn1153Ser	Asn576

B

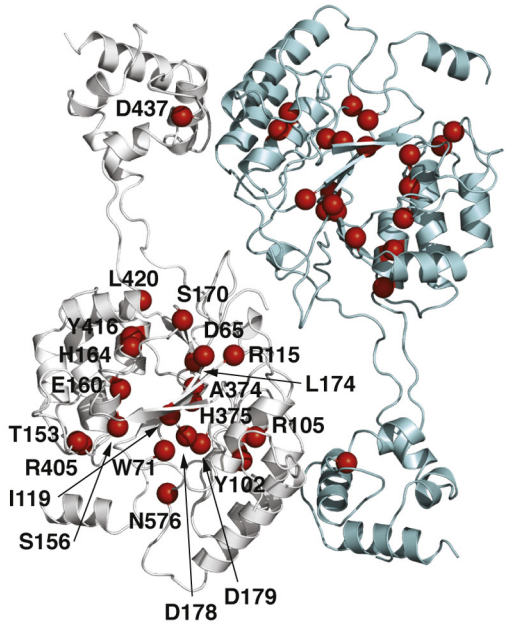

Figure 5. Insights into human mucolipidoses. *A*, select human GNPTAB missense mutations and the corresponding residues in DmGNPTAB. *B*, the positions of DmGNPTAB residues listed in the previous panel are shown in the structure.

In summary, we have elucidated the cryo-EM structure of the *D. melanogaster* GNPTAB homolog, which allows extrapolation of the core structure of human GlcNAc-1-phosphotransferase. One deficiency of our study is the lack of further mechanistic insight into the function of human GlcNAc-1-phosphotransferase. In particular, the most critical issue regarding the molecular mechanism of GlcNAc-1-phosphotransferase, namely, its specific targeting of 60 lysosomal proteins among thousands of glycoproteins that traverse the Golgi apparatus, remains to be addressed. The molecular interaction between the α/β- and γ-subunits also remains to be structurally characterized. Nevertheless, our structural data offer a valuable model to advance the understanding of human GNPTAB and also provide deeper insights into the pathogenesis of human mucolipidoses caused by GNPTAB mutations.

Experimental procedures

Protein expression and purification

The *D. melanogaster GNPTAB* gene (CG8027) was cloned from a fly cDNA library. The DNA fragment encoding DmGNPTAB residues 50 to 630 was cloned into the psMBP2 vector (25), which facilitates its expression in insect cells as a secreted His6-MBP fusion protein with a tobacco etch virus protease cleavage site. Bacmids were generated in DH10Bac cells using the Bac-to-Bac system (Invitrogen). Sf21 insect cells grown in SIM SF medium (Sino Biological Inc) were used to generate and amplify the baculoviruses. For protein production, Hi5 cells grown in SIM HF medium (Sino Biological Inc) were infected at a density of 1.5 to 2.0 × 10^6 cells/ml. Forty-eight hours later, conditioned media were collected by centrifugation at 2000*g* for 30 min. The media were then concentrated using a Hydrosart Ultrafilter (Sartorius) and transferred into binding buffer containing 25 mM Tris-HCl, pH 8.0, and 150 mM NaCl. The recombinant protein was then isolated using Ni-NTA resin (GE Healthcare) and eluted with a buffer containing 25 mM Tris-HCl, pH 8.0, 150 mM NaCl, and 500 mM imidazole. After tobacco etch virus protease digestion for 10 h at 4 °C to remove the N-terminal His6-MBP tag, the protein mixture was transferred into a buffer containing 25 mM Tris-HCl, pH 8.0, and 50 mM NaCl using a Centricon with a cutoff of 10 kDa (Millipore). The untagged proteins were then purified by anion exchange chromatography (Resource Q) and eluted using a 50 to 1000 mM NaCl salt gradient in 25 mM Tris-HCl, pH 8.0, followed by size exclusion chromatography (Superdex Increase 200) and elution in 25 mM Hepes, pH 7.5, and 150 mM NaCl.

Cryo-EM data collection, model building, and structure analyses

Four-microliter aliquots of purified DmGNPTAB at 0.5 mg/ml were applied onto glow-discharged Quantifoil holey-carbon grids (R1.2/1.3, 300 mesh, gold), blotted at 4 °C in 100% humidity, and plunged into liquid ethane with a Vitrobot Mark IV (FEI). The cryogrids were screened with a 200-kV Talos Arctica microscope. Data collection was performed with a 300-kV Titan Krios G3 microscope equipped with a Gatan GIF Quantum K2 Summit direct electron detector using SerialEM

(26). The statistics for data collection and processing are summarized in Table 1.

Movie frames were aligned with MotionCor2 (27). The contrast transfer function parameters were estimated using Gctf (v1.063) (28). Micrographs were sorted based on image qualities, and 14,345 micrographs were used for subsequent reconstruction with Relion (v3.07) (29). Particles were auto-picked based on the templates generated by manual picking and subjected to 2D classification. Classes showing clear structural details were then used to create an initial model. 3D classifications with C1 symmetry were first performed, and 544,196 particles with good structural features were selected for further 3D classifications using C2 symmetry. A total of 131,301 particles were then selected for 3D refinement, which resulted in a map with an overall resolution of 3.53 Å after Bayesian polishing and contrast transfer function refinement. The resolution was estimated using the gold-standard Fourier shell correlation 0.143 criteria.

The structure of DmGNPTAB was built *de novo* in Coot (30). Bulky aromatic residues and N-linked glycosylation sites were used as landmarks during the structural modeling process. Structure refinement was performed using real-space refinement in Phenix (v1.18) (31). A structural homology search was performed using the DALI server (32). Figures were prepared with ESPript (33), PyMOL (Schrödinger), and UCSF Chimera (34).

Generation and characterization of *GNPTAB*$^{-/-}$ HeLa cells

HeLa cells were grown in Dulbecco's modified Eagle's medium supplemented with 10% (v/v) fetal bovine serum (FBS) at 37 °C in a 5% CO_2 incubator. *GNPTAB*$^{-/-}$ cells were generated using CRISPR/Cas9 technology. The guide RNA, 5′- ACAAAACATG GTATTGATCT-3′, which targets exon 2 of *GNPTAB*, was cloned into the pSpCas9(BB)-2A-GFP vector (Addgene, 48,138). Three micrograms of the plasmid was then transfected into HeLa cells using Lipofectamine 2000 (Thermo Scientific). Two days after transfection, GFP-positive cells were sorted into single clones using an Astrios EQ cell sorter (Beckman Coulter). Single clones were cultured in 96-well plates for 2 weeks. The genome type of the knockout cells was determined by DNA sequencing.

For the visualization of lysosomes, parental and GNPTAB$^{-/-}$ HeLa cells were fixed with 4% paraformaldehyde in phosphate-buffered saline (PBS) at room temperature for 15 min, permeabilized with PBS containing 0.1% saponin and 2% bovine serum albumin, and then incubated with anti-LAMP1 antibody (Santa Cruz, sc-20011, 1:50) at 4 °C overnight. The next day, the cells were washed with PBS and incubated with Alexa Fluor 488 donkey anti-mouse antibody (Invitrogen, A21202, 1:150) for 1 h at 25 °C. The coverslips containing the cells were then re-washed with PBS, incubated with DAPI for the visualization of nuclei, and examined with a Delta Vision microscope. The images were analyzed using Velocity (v6.1.1) software.

To examine the maturation of CatB, WT and mutant GNPTAB-Flag plasmids were transfected into *GNPTAB*$^{-/-}$ HeLa cells as indicated. Seventy-two hours later, the cells were harvested, washed with PBS, lysed in lysis buffer (PBS supplemented with 0.5% Triton X-100, 1% protease inhibitor cocktail, and 1 mM phenylmethylsulfonyl fluoride) for 15 min at 4 °C, and analyzed by Western blotting.

Immunoprecipitation and Western blotting

The human *GNPTAB* gene was cloned into modified pcDNA 3.1 vectors that encode a C-terminal Flag or V5 tag. Mutations were generated using a PCR-based method and verified by sequencing. For the immunoprecipitation experiments, V5-tagged GNPTAB was cotransfected with Flag-tagged GNPTAB into HEK293T cells in 10-cm dishes using polyethyleneimine. Forty-eight hours later, the cells were harvested, washed with PBS, and lysed in lysis buffer. The lysates were then centrifuged for 15 min at 13,000*g*, and the supernatant was incubated with Flag M2 beads (Sigma, A2220) for 2 h at 4 °C. The beads were washed three times with lysis buffer. The immunoprecipitated proteins were eluted from the beads using the 3× Flag peptide (NJPeptide, NJP50002) and analyzed by Western blotting.

For the Western blot analyses, protein samples separated by SDS-PAGE were transferred to nitrocellulose membranes, and the membranes were then blocked with 4% nonfat milk for 30 min at 25 °C and incubated with primary antibody overnight at 4 °C. The next day, the membranes were incubated with HRP-conjugated secondary antibodies in 4% nonfat milk for 1 h at room temperature. Detections were performed by enhanced chemiluminescence using an Amersham Imager 800. The primary antibodies used for immunoblotting were anti-Flag (ABclonal, AE005, 1:1000), anti-Flag (MBL, PM020, 1:1000), anti-V5 (Santa Cruz, sc-81594, 1:1000), anti-V5 (Millipore, AB3792, 1:1000), anti-CatB (Cell Signaling, 31,718, 1:1000), and anti-GAPDH (TransGen, HC301-02, 1:5000). The secondary antibodies were goat anti-mouse (TransGen, HS201-01, 1:5000) and goat anti-rabbit (TransGen, HS101-01, 1:5000).

Data availability

The cryo-EM map and atomic coordinates of DmGNPTAB have been deposited in the EMDB and PDB with accession codes EMD-30910 and 7DXI, respectively.

Supporting information—This article contains supporting information.

Acknowledgments—We thank the Core Facilities at the School of Life Sciences, Peking University, for help with negative-staining EM; the Cryo-EM Platform of Peking University for help with data collection; the High-performance Computing Platform of Peking University for help with computation; and the National Center for Protein Sciences at Peking University for assistance with cell sorting.

Author contributions—S. D., G. W., Z. Z., and C. M. investigation; N. G. and J. X. supervision; J. X. conceptualization; J. X. writing-original draft; J. X. funding acquisition.

Funding and additional information—The work was supported by the National Key Research and Development Program of China

(2017YFA0505200 to J. X., 2019YFA0508904 to N. G.), the National Science Foundation of China (31822014 to J. X., 31725007 and 31630087 to N. G.), and the Qidong-SLS Innovation Fund to J. X. and N. G.

Conflict of interest—The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are: CatB, cathepsin B; cryo-EM, cryo-electron microscopy; CR1 to CR4, conserved regions in the Stealth proteins; DMAP, DNA methyltransferase-associated protein interaction domain; M6P, mannose 6-phosphate; N1 and N2, Notch repeats; S1 to S4, spacer regions; TdUGGT, *Thermomyces dupontii* UGGT; UGGT, UDP–glucose glycoprotein glucosyltransferase.

References

1. Varki, A., and Kornfeld, S. (2015) P-type lectins. In: Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Darvill, A. G., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., Seeberger, P. H., eds. *Essentials of Glycobiology*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY: 423–433
2. Bao, M., Booth, J. L., Elmendorf, B. J., and Canfield, W. M. (1996) Bovine UDP-N-acetylglucosamine:lysosomal-enzyme N-acetylglucosamine-1-phosphotransferase. I. Purification and subunit structure. *J. Biol. Chem.* **271**, 31437–31445
3. Kudo, M., and Canfield, W. M. (2006) Structural requirements for efficient processing and activation of recombinant human UDP-N-acetylglucosamine:lysosomal-enzyme-N-acetylglucosamine-1-phosphotransferase. *J. Biol. Chem.* **281**, 11761–11768
4. Sperisen, P., Schmid, C. D., Bucher, P., and Zilian, O. (2005) Stealth proteins: In silico identification of a novel protein family rendering bacterial pathogens invisible to host immune defense. *PLoS Comput. Biol.* **1**, e63
5. Qian, Y., Flanagan-Steet, H., van Meel, E., Steet, R., and Kornfeld, S. A. (2013) The DMAP interaction domain of UDP-GlcNAc:lysosomal enzyme N-acetylglucosamine-1-phosphotransferase is a substrate recognition module. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 10246–10251
6. van Meel, E., Lee, W. S., Liu, L., Qian, Y., Doray, B., and Kornfeld, S. (2016) Multiple domains of GlcNAc-1-phosphotransferase mediate recognition of lysosomal enzymes. *J. Biol. Chem.* **291**, 8295–8307
7. Qian, Y., van Meel, E., Flanagan-Steet, H., Yox, A., Steet, R., and Kornfeld, S. (2015) Analysis of mucolipidosis II/III GNPTAB missense mutations identifies domains of UDP-GlcNAc:lysosomal enzyme GlcNAc-1-phosphotransferase involved in catalytic function and lysosomal enzyme recognition. *J. Biol. Chem.* **290**, 3045–3056
8. De Pace, R., Velho, R. V., Encarnacao, M., Marschner, K., Braulke, T., and Pohl, S. (2015) Subunit interactions of the disease-related hexameric GlcNAc-1-phosphotransferase complex. *Hum. Mol. Genet.* **24**, 6826–6835
9. Velho, R. V., De Pace, R., Tidow, H., Braulke, T., and Pohl, S. (2016) Identification of the interaction domains between alpha- and gamma-subunits of GlcNAc-1-phosphotransferase. *FEBS Lett.* **590**, 4287–4295
10. Marschner, K., Kollmann, K., Schweizer, M., Braulke, T., and Pohl, S. (2011) A key enzyme in the biogenesis of lysosomes is a protease that regulates cholesterol metabolism. *Science* **333**, 87–90
11. Liu, L., Lee, W. S., Doray, B., and Kornfeld, S. (2017) Role of spacer-1 in the maturation and function of GlcNAc-1-phosphotransferase. *FEBS Lett.* **591**, 47–55
12. Qian, Y., Lee, I., Lee, W. S., Qian, M., Kudo, M., Canfield, W. M., Lobel, P., and Kornfeld, S. (2010) Functions of the alpha, beta, and gamma subunits of UDP-GlcNAc:lysosomal enzyme N-acetylglucosamine-1-phosphotransferase. *J. Biol. Chem.* **285**, 3360–3370
13. Velho, R. V., Harms, F. L., Danyukova, T., Ludwig, N. F., Friez, M. J., Cathey, S. S., Filocamo, M., Tappino, B., Gunes, N., Tuysuz, B., Tylee, K. L., Brammeier, K. L., Heptinstall, L., Oussoren, E., van der Ploeg, A. T., *et al.* (2019) The lysosomal storage disorders mucolipidosis type II, type III alpha/beta, and type III gamma: Update on GNPTAB and GNPTG mutations. *Hum. Mutat.* **40**, 842–864
14. Frigerio-Domingues, C., and Drayna, D. (2017) Genetic contributions to stuttering: The current evidence. *Mol. Genet. Genomic Med.* **5**, 95–102
15. Flint, M., Chatterjee, P., Lin, D. L., McMullan, L. K., Shrivastava-Ranjan, P., Bergeron, E., Lo, M. K., Welch, S. R., Nichol, S. T., Tai, A. W., and Spiropoulou, C. F. (2019) A genome-wide CRISPR screen identifies N-acetylglucosamine-1-phosphate transferase as a potential antiviral target for Ebola virus. *Nat. Commun.* **10**, 285
16. Wang, R., Simoneau, C. R., Kulsuptrakul, J., Bouhaddou, M., Travisano, K. A., Hayashi, J. M., Carlson-Stevermer, J., Zengel, J. R., Richards, C. M., Fozouni, P., Oki, J., Rodriguez, L., Joehnk, B., Walcott, K., Holden, K., *et al.* (2021) Genetic screens identify host factors for SARS-CoV-2 and common cold coronaviruses. *Cell* **184**, 106–119.e14
17. Tiede, S., Storch, S., Lubke, T., Henrissat, B., Bargal, R., Raas-Rothschild, A., and Braulke, T. (2005) Mucolipidosis II is caused by mutations in GNPTA encoding the alpha/beta GlcNAc-1-phosphotransferase. *Nat. Med.* **11**, 1109–1112
18. Lairson, L. L., Henrissat, B., Davies, G. J., and Withers, S. G. (2008) Glycosyltransferases: Structures, functions, and mechanisms. *Annu. Rev. Biochem.* **77**, 521–555
19. Braakman, I., and Bulleid, N. J. (2011) Protein folding and modification in the mammalian endoplasmic reticulum. *Annu. Rev. Biochem.* **80**, 71–99
20. Roversi, P., Marti, L., Caputo, A. T., Alonzi, D. S., Hill, J. C., Dent, K. C., Kumar, A., Levasseur, M. D., Lia, A., Waksman, T., Basu, S., Soto Albrecht, Y., Qian, K., McIvor, J. P., Lipp, C. B., *et al.* (2017) Interdomain conformational flexibility underpins the activity of UGGT, the eukaryotic glycoprotein secretion checkpoint. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 8544–8549
21. Satoh, T., Song, C., Zhu, T., Toshimori, T., Murata, K., Hayashi, Y., Kamikubo, H., Uchihashi, T., and Kato, K. (2017) Visualisation of a flexible modular structure of the ER folding-sensor enzyme UGGT. *Sci. Rep.* **7**, 12142
22. Encarnacao, M., Kollmann, K., Trusch, M., Braulke, T., and Pohl, S. (2011) Post-translational modifications of the gamma-subunit affect intracellular trafficking and complex assembly of GlcNAc-1-phosphotransferase. *J. Biol. Chem.* **286**, 5311–5318
23. Ludwig, N. F., Velho, R. V., Sperb-Ludwig, F., Acosta, A. X., Ribeiro, E. M., Kim, C. A., Gandelman Horovitz, D. D., Boy, R., Rodovalho-Doriqui, M. J., Lourenco, C. M., Santos, E. S., Braulke, T., Pohl, S., and Schwartz, I. V. D. (2017) GNPTAB missense mutations cause loss of GlcNAc-1-phosphotransferase activity in mucolipidosis type II through distinct mechanisms. *Int. J. Biochem. Cell Biol.* **92**, 90–94
24. Danyukova, T., Ludwig, N. F., Velho, R. V., Harms, F. L., Gunes, N., Tidow, H., Schwartz, I. V., Tuysuz, B., and Pohl, S. (2020) Combined *in vitro* and in silico analyses of missense mutations in GNPTAB provide new insights into the molecular bases of mucolipidosis II and III alpha/beta. *Hum. Mutat.* **41**, 133–139
25. Tagliabracci, V. S., Wen, J., and Xiao, J. (2016) Methods to purify and assay secretory pathway kinases. *Methods Mol. Biol.* **1496**, 197–215
26. Mastronarde, D. N. (2005) Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51
27. Zheng, S. Q., Palovcak, E., Armache, J. P., Verba, K. A., Cheng, Y., and Agard, D. A. (2017) MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332
28. Zhang, K. (2016) Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12
29. Zivanov, J., Nakane, T., Forsberg, B. O., Kimanius, D., Hagen, W. J., Lindahl, E., and Scheres, S. H. (2018) New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, e42166
30. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501
31. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R.

W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., *et al.* (2010) PHENIX: A comprehensive python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221

32. Holm, L. (2020) DALI and the persistence of protein shape. *Protein Sci.* **29**, 128–140
33. Gouet, P., Robert, X., and Courcelle, E. (2003) ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res.* **31**, 3320–3323
34. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612