PNAS RESEARCH ARTICLE | BIOCHEMISTRY

Structures of the mannose-6-phosphate pathway enzyme, GlcNAc-1-phosphotransferase

Alexei Gorelik[a,b], Katalin Illes[a], Khanh Huy Bui[b], and Bhushan Nagar[a,1]

Edited by Robert Stroud, University of California, San Francisco, California; received February 26, 2022; accepted May 6, 2022

The mannose-6-phosphate (M6P) pathway is responsible for the transport of hydrolytic enzymes to lysosomes. N-acetylglucosamine-1-phosphotransferase (GNPT) catalyzes the first step of tagging these hydrolases with M6P, which when recognized by receptors in the Golgi diverts them to lysosomes. Genetic defects in the GNPT subunits, GNPTAB and GNPTG, cause the lysosomal storage diseases mucolipidosis types II and III. To better understand its function, we determined partial three-dimensional structures of the GNPT complex. The catalytic domain contains a deep cavity for binding of uridine diphosphate-*N*-acetylglucosamine, and the surrounding residues point to a one-step transfer mechanism. An isolated structure of the gamma subunit of GNPT reveals that it can bind to mannose-containing glycans in different configurations, suggesting that it may play a role in directing glycans into the active site. These findings may facilitate the development of therapies for lysosomal storage diseases.

GlcNAc-1-phosphotransferase | mannose-6-phosphate | lysosome

Significance

Lysosomes are cellular organelles containing a variety of degradative enzymes. The protein N-acetylglucosamine-1-phosphotransferase (GNPT) is responsible for directing these enzymes to be delivered to lysosomes. Genetic mutations in GNPT cause the diseases mucolipidosis types II and III. In order to better understand the function of GNPT at the molecular level, we determined the three-dimensional structures of its catalytic domain and of its sugar-binding subunit. These structures, along with biochemical experiments, reveal its enzymatic mechanism and its sugar recognition mode.

N-linked glycosylation is a common modification of eukaryotic proteins that transit through the secretory pathway. Nascent polypeptides containing a signal sequence are directed into the endoplasmic reticulum, where most are N-glycosylated. Their glycans undergo processing in the Golgi, followed by transport in secretory vesicles to the cell surface. A small subset of proteins is instead directed to late endosomes and ultimately to lysosomes. Out of the ~2,700 human soluble N-glycosylated proteins (1), around 70 are lysosomal hydrolytic enzymes that catabolize proteins, polysaccharides, lipids, and other molecules (2).

Transport of soluble enzymes to lysosomes mainly occurs via the mannose-6-phosphate (M6P) pathway (3, 4). N-linked glycans generally contain multiple mannose units, and the glycans of proteins destined to lysosomes specifically acquire a phosphate group on a mannose residue. This tag is bound by two transmembrane receptors in the *trans*-Golgi, CD-MPR and CI-MPR, leading to the inclusion of the tagged protein in transport vesicles (5). The M6P marker is formed in two steps. First, the *cis*-Golgi transmembrane enzyme N-acetylglucosamine-1-phosphotransferase (GNPT) recognizes transiting proteins that must be directed to lysosomes and transfers a phospho-*N*-acetylglucosamine group onto their glycan, resulting in a sugar–phosphate–sugar intermediate (6–8). The terminal *N*-acetylglucosamine sugar is then removed by a second enzyme, *N*-acetylglucosamine-1-phosphodiester alpha-*N*-acetylglucosaminidase (NAGPA), also called "uncovering enzyme" (9–11).

GNPT, the first enzyme of the M6P pathway, is a 400-kDa hexameric complex of three types of subunits, $\alpha_2\beta_2\gamma_2$ (12). The α and β subunits are generated by proteolytic activation of the precursor protein GNPTAB (13–15) and contain the catalytic domain (13, 16) as well as additional modules that recognize and bind to lysosomal hydrolases (17–19). The γ subunit (GNPTG) comprises a mannose-binding domain and also contributes to substrate selection (19, 20); it is required for proper transport of approximately two-thirds of lysosomal hydrolases (2, 20, 21).

Genetic mutations in GNPTAB and GNPTG cause the lysosomal storage diseases mucolipidosis type II, fatal in early childhood, and type III, a less severe form (16, 22–24). There are currently no disease-specific treatments for these disorders (25), with preclinical investigations of gene therapy (26) and antisense oligonucleotides (27) in progress. These approaches, and others including pharmacological chaperones (28), could be facilitated by detailed structural information on GNPT. A major unanswered question is its mechanism of specific recognition of 70 lysosomal hydrolases, which share no structural or sequence similarity among themselves, from the 2,700 soluble glycoproteins of the secretory pathway (8, 29). This process requires features at specific positions on the surface of these enzymes involving lysine residues (30–35), but the details of these interactions are unknown. GNPT is not homologous to other vertebrate proteins, and although it was discovered over 40 years ago (7) its three-dimensional structure is mostly uncharacterized. Another

Author affiliations: [a]Department of Biochemistry, McGill University, Montreal, QC H3G 0B1, Canada; and [b]Department of Anatomy and Cell Biology, McGill University, Montreal, QC H3A 0C7, Canada

Author contributions: A.G. and K.I. designed research; A.G. and K.I. performed research; A.G. analyzed data; and A.G., K.I., K.H.B., and B.N. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: bhushan.nagar@mcgill.ca.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2203518119/-/DCSupplemental.

Published August 8, 2022.

unclear aspect is its catalytic mechanism, as it carries out a reaction different from related enzymes of known structure.

Here, we report a partial characterization of GNPT by crystallography. Four segments dispersed across the primary sequence assemble into a compact catalytic domain. The donor substrate uridine diphosphate (UDP)-*N*-acetylglucosamine is tightly bound in the active site cavity, exposing one phosphate group. A reaction mechanism is proposed, supported by in vitro activity assays. The glycan-binding mode of the isolated γ subunit is also determined. These results significantly increase our understanding of the function of this unique enzyme.

During the preparation of this paper, the three-dimensional structure of a drosophila GNPT homolog was reported (36). This protein lacks the γ subunits and most of the noncatalytic domains, and its physiological function is unknown; nevertheless, it shares the same fold as the structure described here.

Results and Discussion

Crystal Structure of the GNPT Catalytic Domain. GNPT is a 400-kDa heterohexamer of three types of subunits, $\alpha_2\beta_2\gamma_2$ (12). The α and β proteins are proteolytic products (14, 15) encoded by a single gene (13, 16). The complex is anchored to the *cis*-Golgi membrane (6) by transmembrane helices at both ends of the GNPTAB precursor, with short cytoplasmic overhangs (Fig. 1*A*). High-resolution characterization of the soluble luminal portion of $\alpha_2\beta_2$ or $\alpha_2\beta_2\gamma_2$ complexes by crystallography and electron microscopy was unsuccessful, so the protein's catalytic domain was investigated next.

The GNPT enzymatic activity is carried out by the α and β subunits (37), whereas γ is auxiliary (20). The catalytic domain is composed of four segments dispersed across the GNPTAB polypeptide, separated by three large inserted regions (Fig. 1*A*) (17), as deduced from sequence similarity to bacterial enzymes of the Stealth family (13, 16, 38). A "minimal" GNPT catalytic construct of α and β—more suitable for structural studies—in which the first and second insertions were removed was previously reported (39). The first of these deleted regions comprises two domains structurally similar to RNA recognition motifs (RRM) (40); they are the only GNPT portion of known three-dimensional structure (Protein Data Bank [PDB] ID 2N6D), but their function is unclear (41). The second removed segment

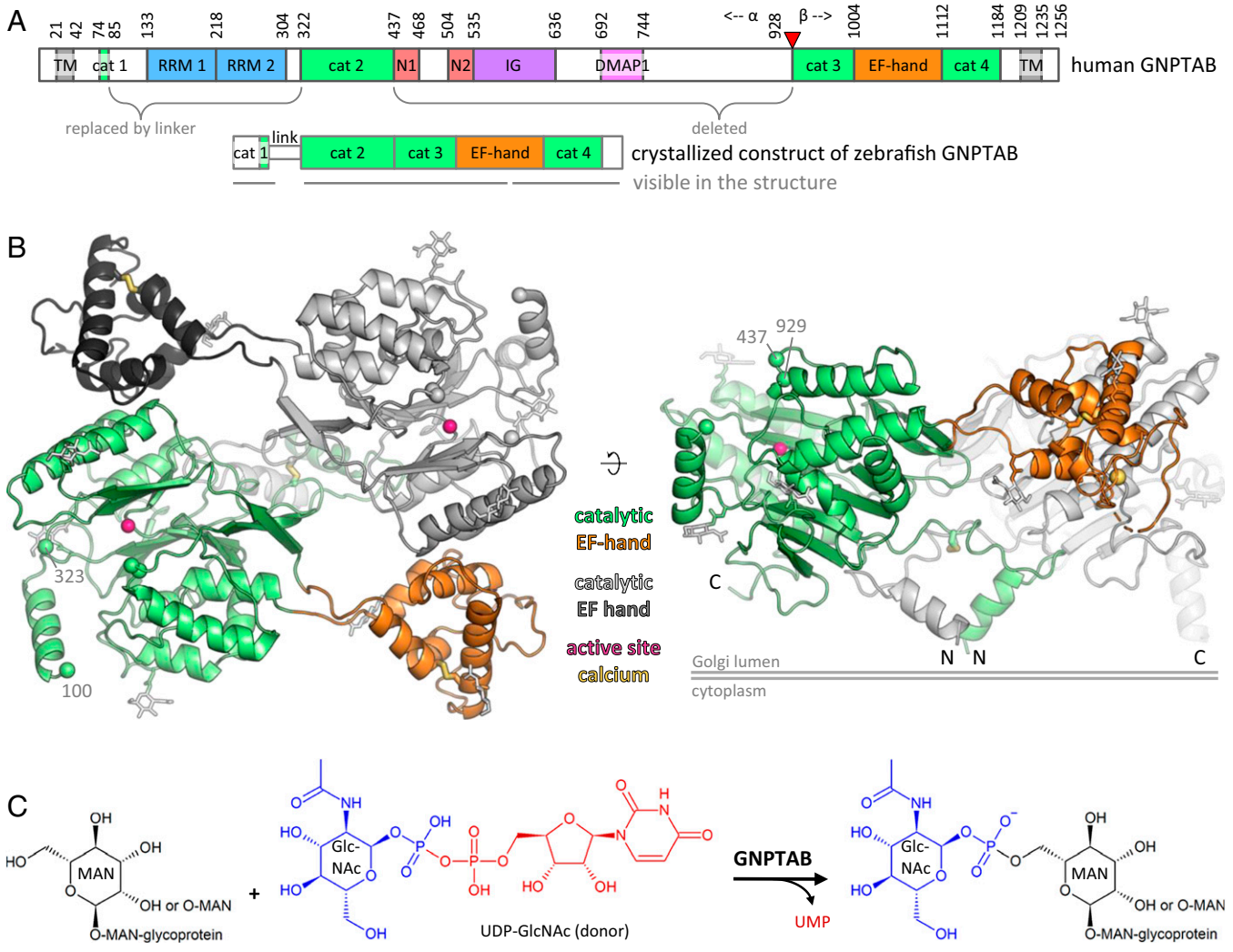


Fig. 1. Structure of the GNPTAB catalytic domain. (*A*) Domain organization of vertebrate GNPTAB. TM, transmembrane helix; cat, catalytic domain segment; RRM, RNA recognition motif-like domain; N, Notch extracellular repeats-like domain; IG, immunoglobulin domain; DMAP1, DMAP1 binding-like domain; EF-hand, calcium-binding domain. The site-1 protease (S1P) cleavage site between the α and β subunits is marked by a red triangle. (*B*) Crystal structure of the zebrafish minimal construct dimer, comprising the catalytic (green, light gray) and EF-hand (orange, dark gray) domains. Calcium ions are represented by yellow spheres and the active sites by pink spheres. The N and C termini are labeled. Disulfide bonds are shown as thick yellow sticks. N-linked glycans (white sticks) are simplified for clarity. The deletions in the minimal construct are labeled (human residue numbers) and marked by green spheres. (*C*) Phosphotransfer reaction catalyzed by GNPTAB. MAN, mannose; GlcNAc, *N*-acetylglucosamine.

is composed of two Notch-like (EGF-like) repeats (13, 16, 42), a predicted immunoglobulin module (43), a domain similar to the DMAP1-binding region of several proteins (16, 17), and a 180-residue-long segment predicted to be unstructured. These domains participate in specific recognition and binding to lysosomal hydrolases, allowing the catalytic domain to tag their glycans (17–19). The minimal GNPT construct, lacking these accessory domains, remains catalytically active but cannot discriminate between lysosomal hydrolases and nonlysosomal glycoproteins (39). It also lacks the binding site for the γ subunit (18, 19, 44).

We determined the crystal structure of the GNPT minimal construct from zebrafish (87% sequence identity to human) (Fig. 1*B*). The crystallized protein comprises the four portions of the catalytic domain, as well as a small domain with an EF-hand motif (Fig. 1*A*) (45). At the core of the catalytic domain are three roughly parallel β-sheets containing five, three, and two strands, flanked by three α-helices on one side and five on the other. The four conserved segments, distant from each other in the GNPTAB primary sequence, all contribute to the central β-sheets (*SI Appendix*, Fig. S1*A*), so both the α and β subunits are required for activity. This domain is part of a structural superfamily (CATH ID 3.90.550.10) containing sugar- and nucleotide-transferases but shares less than 15% sequence identity with proteins of known structure or with any other vertebrate enzymes. It is most similar to the Stealth family of bacterial capsular polysaccharide synthesis enzymes (38), which have not been structurally characterized and which lack the ancillary modules found in GNPT. The EF-hand domain projects sideways from the protein via two 13-Å linkers and consists of a five-helix bundle stabilized by a disulfide bond (human residues 1057 to 1094) and by a bound calcium ion. Up to five N-linked glycans are visible in the structure on each protein chain, out of seven predicted sites.

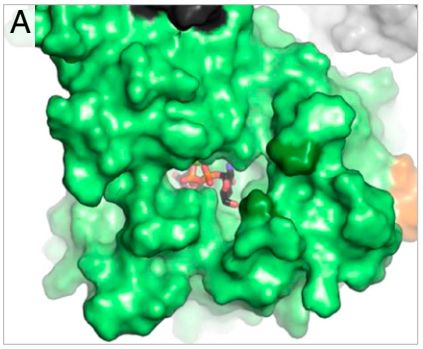
The native GNPT complex contains two copies of the α, β, and γ subunits (12), and accordingly the minimal construct crystallized as a dimer (Fig. 1*B*). The 2,650-Å^2 dimerization interface is contributed by mostly charged and polar interactions between each EF-hand domain and the opposing catalytic domain (1,150 Å^2), as well as polar and hydrophobic contacts between both catalytic domains (1,500 Å^2). The latter are mainly mediated by an N-terminal segment swapped between both subunits and stabilized by an intermolecular disulfide bond (Cys70) (12, 44). This arrangement very likely corresponds to the physiological dimerization within the GNPT complex, which is anchored to the Golgi membrane by four transmembrane helices—one at each end of the GNPTAB precursor (13). The N and C termini of both chains in the minimal construct crystal structure are found on the same side (Fig. 1*B*), suggesting a membrane-facing orientation for that side of the protein. The two putative active sites of the dimer are located 60 Å apart, in a surface cavity facing away from the membrane, into the lumen.

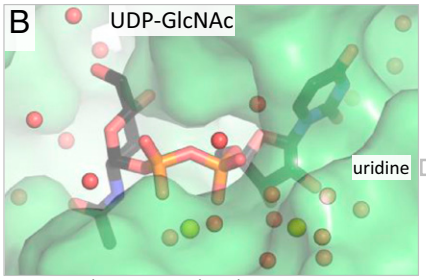
Enzymatic Mechanism of GNPT. GNPT carries out the first step of mannose-6-phosphate tag formation, namely the transfer of phospho-*N*-acetylglucosamine (phospho-GlcNAc) to a mannose residue part of the N-linked glycan on substrate proteins. This group is transferred from UDP-GlcNAc, a nucleotide-sugar that serves as donor for certain glycosyltransferases (Fig. 1*C*). The GNPT catalytic domain is part of a structural superfamily containing sugar- and nucleotide-transferases (*SI Appendix*, Fig. S1*B*) (46, 47). However, it catalyzes a phosphotransfer—not a sugar transfer—so its mechanism must be different from that of related proteins which have been structurally characterized. To investigate it, we determined the crystal structure of the zebrafish GNPT minimal construct in complex with the donor substrate UDP-GlcNAc. The molecule occupies a cavity on the catalytic domain facing away from the membrane (Fig. 2*A*). The opening of this cavity is quite narrow, suggesting that minor rearrangements take place to allow donor entry or product (UMP) exit. The location of the binding site is conserved in structurally related enzymes (*SI Appendix*, Fig. S1*C*). The high resolution (2.3 Å) of the structure allows visualization of the multiple contacts, including ones mediated by water molecules, between the protein and different parts of the donor substrate (Fig. 2*B* and *SI Appendix*, Fig. S2*A*).

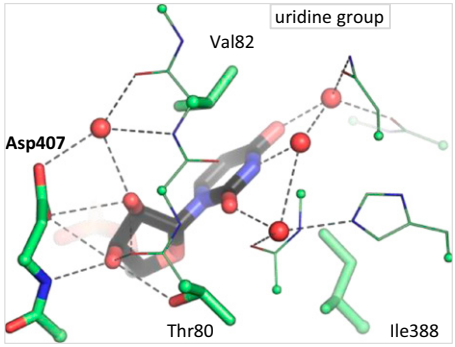
The uridine moiety of UDP-GlcNAc is positioned deep in the cavity, with its nucleobase sandwiched between the hydrophobic side chains of Val82 and Ile388. There does not seem to be strong base selectivity, as uracil interacts with the protein indirectly via three H-bonded water molecules. The ribose 2′ and 3′ hydroxyl groups form five H-bonds, notably with Asp407 (human residue numbering). At the other end of the cavity, the GlcNAc moiety is held by 12 H-bonds, including five water-mediated ones, and directly contacts Asn406, Glu389, Gln992, His956, His959, and Ser385. To assess the functions of several of these residues in a more physiological construct, mutations were introduced in the hamster soluble (luminal portion) active $\alpha_2\beta_2\gamma_2$ complex, and its ability to tag the small-molecule substrate α-methyl D-mannoside (Fig. 2*C*) or the glycans of lysosomal acid ceramidase (Fig. 2*F* and *SI Appendix*, Fig. S2*E*) was assayed in vitro. Disruption of Glu389, Gln992, Asp407, or His956 decreases activity by 50- to 1,000-fold, although the last two residues serve additional functions described below.

The phosphate group attached to GlcNAc (β) is the most solvent-accessible portion of the substrate, whereas the uridine-linked phosphate (α) is more obstructed (Fig. 2*A*). GNPT requires magnesium or manganese for activity (29, 37), and indeed a magnesium ion (Mg1) is coordinated by both phosphate groups at once, as well as by Asp408 and Asn1151 (Fig. 2*B*). Glycosyltransferases related to GNPT often contain a divalent ion in their active site, positioned similarly to Mg1 (48). The β phosphate, which will be transferred onto the substrate glycan, is further held in place by two salt bridges with Arg986 and H-bonds with Asn1151 and three water molecules. The α phosphate, which constitutes the leaving group in the reaction, interacts with several water molecules. Unexpectedly, a second magnesium ion (Mg2) is found in the active site. It does not contact the protein directly, being fully coordinated by six water molecules—two of which interact with the α phosphate, two with Asp407, and all of them involved in an extensive H-bonding network with multiple side chains and main chain atoms. In fact, the presence of Mg2 stabilizes a segment that is disordered in the apo protein (*SI Appendix*, Fig. S2*B*). Mutation of Asp407 decreases activity 500-fold (Fig. 2 *C* and *F*), suggesting that the second magnesium ion may be important for maintaining the active-site architecture; however, this residue also interacts with the ribose moiety as described above.

Although the UDP-GlcNAc–binding site is similarly located in GNPT and in related glycosyltransferases (*SI Appendix*, Fig. S1*C*), they catalyze a different reaction (*SI Appendix*, Fig. S1*B*) and hence use a different mechanism. GNPT is closest to the Stealth family of bacterial enzymes (38), which also transfer phospho-sugars but have not been structurally characterized. The current data allow us to propose a mechanism for this reaction (Fig. 2*E*). It very likely takes place in one step, without a covalent enzyme intermediate, as there are no Glu, Asp, Ser, Thr, Tyr, or Cys side chains within suitable distance of the β phosphate group. The mannose 6-hydroxyl carries out a nucleophilic attack

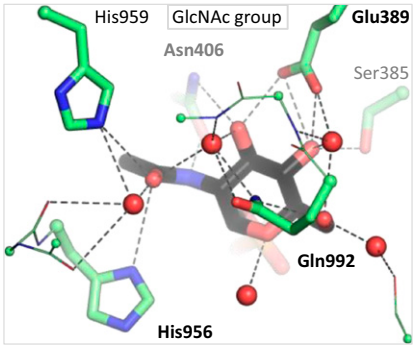


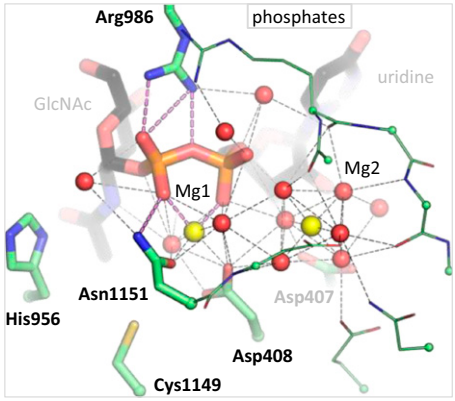


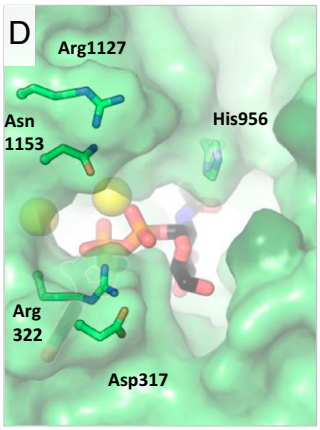


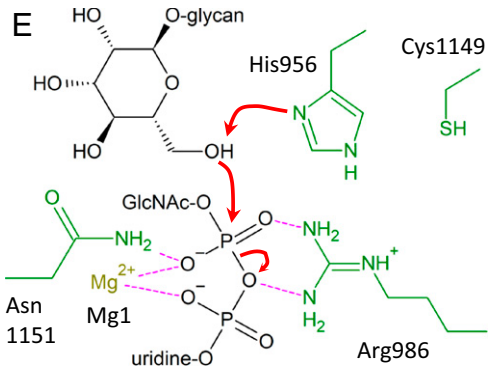
C

hamster GNPT $\alpha_2\beta_2\gamma_2$ activated	relative activity, αMeMAN (%)	conservation in Stealth family (%)
wild-type	100 ± 2.5	
precursor	1.2 ± 0.0	
His956Gln	0.0 ± 0.0	99
His956Asn	0.2 ± 0.0	
Arg986Lys	0.0 ± 0.0	99
Arg986Gln	0.0 ± 0.0	
Asn1151Ala	0.5 ± 0.0	95
Asp408Ala	0.0 ± 0.0	100
Cys1149Ser	11.1 ± 0.3	99
Cys1149Ala	36.5 ± 0.7	
Asp407Ala	0.2 ± 0.0	100
Gln992Leu	1.7 ± 0.0	low
Glu389Leu	0.0 ± 0.0	93
Asn406Ala	5.7 ± 0.2	100
His959		71
Ser385		100
Arg1127Ala	99.2 ± 0.8	low
Asn1153Ala	0.6 ± 0.0	low
Arg322Gln	4.7 ± 0.1	99
Asp317Ala	169.3 ± 5.5	low









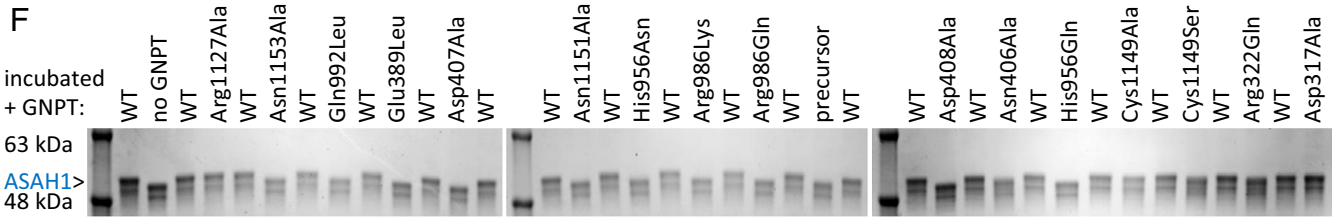


Fig. 2. Structure of the GNPTAB catalytic domain with UDP-GlcNAc. (*A*) GNPT catalytic domain (light green) bound to UDP-GlcNAc (sticks). The two residues in dark green represent the junction point between segments 2 and 3 of the catalytic domain; the full-length protein contains a large insertion there that was removed in the minimal construct. (*B*) Polar and charged contacts (dashed lines) between the protein, UDP-GlcNAc (black sticks), magnesium ions (yellow spheres), and water molecules (red spheres). Distances are below 3.7 Å. Due to the complexity of the zebrafish minimal construct residue numbering and its high sequence identity to the human homolog (87%), human residue numbers are shown throughout this figure. Side chains that were functionally investigated are in bold. (*C*) In vitro enzymatic activity of purified hamster soluble GNPTAB–GNPTG active complex (furin-cleaved, referred to as WT) on the small molecule substrate α-methyl D-mannoside (100 mM) in presence of UDP-GlcNAc (1 mM). The precursor is not furin-cleaved. Values are the means and SDs of a representative of two experiments, each performed in triplicates. The 100% activity corresponds to 0.35/s. Sequence conservation in the 1,000 closest homologs of the GNPT catalytic domain (Stealth family) from unicellular organisms is indicated. (*D*) Residues lining the opening of the active site cavity and potentially interacting with the glycan substrate. (*E*) Proposed catalytic mechanism. (*F*) In vitro activity of GNPT as above on the purified lysosomal hydrolase acid ceramidase (ASAH1), a substrate of GNPT (2, 21). This protein migrates as a smeared double band on the gel due to glycan heterogeneity from recombinant expression. Tagging of its glycans by GNPT shifts the bands to a higher molecular weight.

on that phosphate, which is stabilized and neutralized by the positively charged Arg986 and Mg1 and by Asn1151. The leaving α phosphate group is also stabilized by Arg986 and Mg1. Mutation of Asn1151 decreased activity 200-fold, and the conservative substitution Arg986Lys abolished the reaction (Fig. 2 *C* and *F*). Nucleophilic attack is likely enabled by deprotonation of the mannose 6-hydroxyl by His956 (Fig. 2*B*)—the only residue suitably located for this purpose; the His956Asn/Gln mutants also severely impaired activity, while preserving the His956 H-bond to the GlcNAc moiety. Arg956 and His986 are strictly conserved

in the Stealth family (Fig. 2*C* and *SI Appendix*, Fig. S5), as is Asp408, which coordinates Mg1, further supporting this mechanism. The conservation and functional importance of these residues was previously established (46).

An alternative mechanism involving a covalent intermediate between His956 and the β phosphate is unlikely due to the low but measurable activity of the His956Asn mutant, the over-4-Å distance between the residue and the phosphorus atom even after side-chain rotation, and the absence of such an intermediate in the crystal structure after a prolonged incubation of the crystal with UDP-GlcNAc.

Genetic mutations in GNPTAB cause mucolipidosis type II or type III α/β (16, 22, 23). Notably, the variants Ser385-Leu, Glu389Lys(II), Asp407Ala(III), Asp408Asn(II), His956-Tyr/Arg(III), and Arg986Cys/Gly/His(II) occur in mucolipidosis patients (25); the current study implicates these residues in UDP-GlcNAc binding or catalysis.

GNPTAB is produced as an enzymatically inactive precursor and activated by cleavage after Lys928 by site-1 protease (13–15). The minimal catalytic construct used in this study is missing the segment 438 to 928 (Fig. 1*A*) and does not require proteolytic activation (39). Therefore, this structure does not reveal the activation mechanism or the mode of inhibition in the precursor. The point of fusion between residues 437 and 929 is found near the active site, above the GlcNAc-binding subpocket (*SI Appendix*, Fig. S2*B*). This segment is relatively flexible in the apo protein and disordered in the UDP-GlcNAc complex. One possibility is that the precursor adopts a conformation that prevents efficient substrate entry.

Native GNPT and the minimal construct can tag substrates as small as a single mannose molecule, albeit at low affinity (37, 39). However, no substrate was visible in the structures of crystals incubated with high concentrations of mono- or di-mannose compounds. This is likely due to the very low affinity for these acceptors (5 to 100 mM); in contrast, the Michaelis constant for glycosylated lysosomal hydrolases is in the range of 20 μM (14, 21, 37). In the absence of structural data, the area above the UDP-GlcNAc–binding cavity was examined for side chains that could interact with the terminal (acceptor) or preterminal mannose residues of the glycoprotein substrate (Fig. 2*D*). Substitution of Arg322 or Asn1153 reduced activity on the small-molecule substrate by at least 20-fold and decreased tagging of acid ceramidase (Fig. 2 *C* and *F*), suggesting that these two residues help position the terminal mannose for deprotonation by His956 and reaction with UDP-GlcNAc, as described above. Notably, the Asn1153Ser mutation occurs in mucolipidosis III (25).

The lack of tight interactions with the acceptor sugar substrate was expected, as these glycans are indistinguishable between lysosomal hydrolases and other glycoproteins in the secretory pathway that are not GNPT substrates. Most of the affinity seems to be provided by nonglycan features of the substrate proteins recognized by noncatalytic domains of GNPT or by its γ subunit.

Crystal Structure of GNPTG Bound to Glycans. The γ subunits of the GNPT $\alpha_2\beta_2\gamma_2$ complex are not required for catalysis and for tagging of certain lysosomal hydrolases but are necessary for activity on others (2, 20, 21). This subunit also enhances the addition of a second phospho-GlcNAc group to singly tagged glycans (21). GNPTG contains a signal peptide (24), a mannose-6-phosphate receptor homology (MRH) domain (49, 50), and, similarly to GNPTAB, a DMAP1-interaction domain (Fig. 3*A*) (19). The MRH domain binds to glycans of substrate proteins (19) and was proposed to direct the glycan into the active site of the GNPT complex (21). Meanwhile, the DMAP1-interaction module can specifically recognize features and bind to lysosomal hydrolases (19).

We determined the crystal structure of GNPTG from clawed frog (60% sequence identity to human) (Fig. 3*C*). The crystallized construct encompasses the entire protein; however, a 23-residue N-terminal segment and an 85-residue C-terminal portion outside of the two domains mentioned above are not visible in the structure (Fig. 3*A*). The MRH domain forms the bulk of the protein and consists of a flattened β-barrel composed of nine β-strands, stabilized by three disulfide bonds. Toward one end of the barrel, a lateral hydrophobic patch establishes an interface with the DMAP1-interaction module, which consists of two antiparallel α-helices. A cavity at the opposite end of the barrel forms the mannose-binding site.

The GNPTG MRH domain is part of a family of mannose-binding lectins that includes the mannose-6-phosphate receptors and three proteins involved in quality control in the secretory pathway (*SI Appendix*, Fig. S3*C*) (50) that display up to 30% sequence identity with GNPTG. These domains share a binding site that recognizes glycans with various specificities. In the γ subunit structure, this site is fortuitously occupied by an N-linked glycan attached to a neighboring protein copy from the crystal lattice (Fig. 3*C* and *SI Appendix*, Fig. S3*B*). GNPTG crystallized as a dimer, and the bound glycans extend from different directions into the binding site of both chains (Fig. 3*B*). Due to heterogeneity of the recombinant GNPTG glycan composition, and the crystal lattice arrangement, one γ subunit is bound to an α-1,2 mannose residue and the other to an α-1,3 mannose (Fig. 3*D* and *SI Appendix*, Fig. S3*A*). In both cases, the terminal mannose is held by up to nine hydrogen bonds from four side chains (human residues Gln91, Arg134, Glu153, and Tyr159) conserved in the MRH family (*SI Appendix*, Fig. S3*D*). These residues are required for the integrity of the sugar-binding site and proper function (19). Tyr79 also establishes two hydrogen bonds with the second-to-last mannose in both γ chains; in addition, this mannose residue forms contacts with Glu93 and Arg97, but only in the α-1,2 configuration (Fig. 3*D*). Finally, Asn99 interacts with a mannose from another arm of the glycan, although this could be induced by the crystal lattice arrangement.

GNPT can tag N-linked glycans at five different positions on the mannose tree (*SI Appendix*, Fig. S3*A*) (29, 61), with up to two phospho-GlcNAc tags per glycan (62). The ability of the γ subunit to bind to α-1,2 as well as α-1,3 terminal mannose residues, as seen in the current structure, may be relevant for its proposed mode of action—directing glycans into the active site (21). As the high-mannose glycan is trimmed to a variable extent by α-mannosidases in the *cis*-Golgi where GNPT is located (63, 64), and because GNPT recognizes target proteins of various sizes glycosylated at diverse locations on their surface [sometimes tagging multiple glycans per protein (65)], this binding versatility may be required to recognize glycans of different lengths and orientations. Notably, GNPT facilitates the second tagging of singly tagged glycans (21), possibly by interacting with one arm of the glycan in order to properly position another arm for catalysis. This raises the possibility that GNPTG can accept an already-tagged mannose bearing a phospho-GlcNAc group into its binding pocket. In the MRH family, certain proteins are specific for unmodified mannose (58, 60) whereas others recognize mannose-6-phosphate (51, 53, 57) or even mannose-6-phospho-GlcNAc (55). These differences are mainly mediated by a loop (Fig. 3*E*) (loop 1, residues 129 to 132), the orientation and sequence of which provide specific contacts or steric restrictions. In GNPTG,

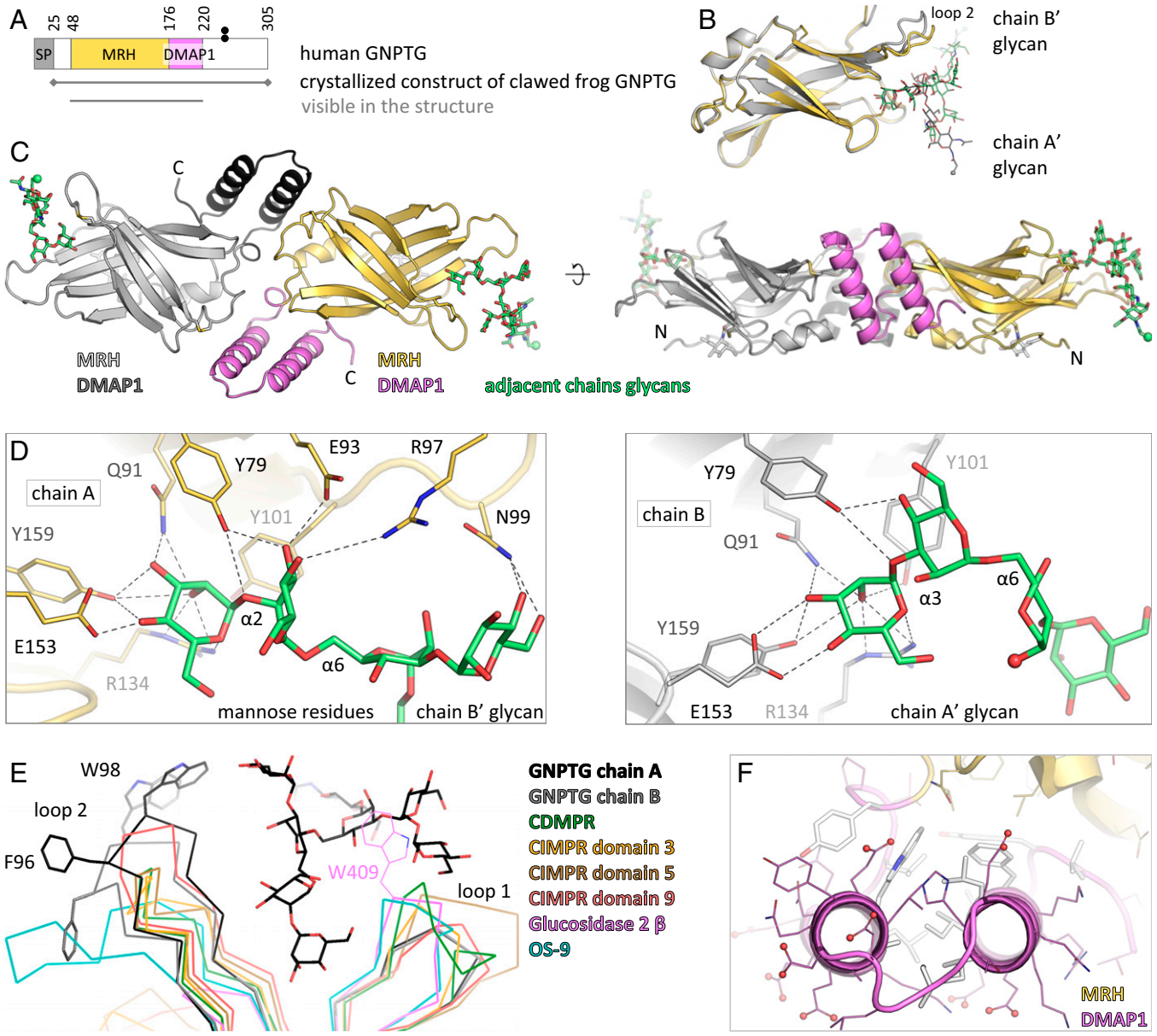


Fig. 3. Structure of GNPTG and its complex with glycans. (*A*) Domain organization of GNPTG. SP, signal peptide; MRH, mannose 6-phosphate receptor homology domain; DMAP1, DMAP1 binding-like domain. The disulfide bond between two GNPTG subunits is represented by black dots. (*B*) The two subunits of the crystallized GNPTG dimer are superimposed, illustrating the different orientations of their bound glycans. A loop that adopts a different angle in chain B is labeled. (*C*) Crystal structure of the clawed frog GNPTG dimer, comprising the MRH (yellow, light gray) and DMAP1 (pink, dark gray) domains. The MRH domain carbohydrate-binding site is occupied by N-linked glycans (green sticks) from adjacent proteins in the crystal lattice (chains B′ and A′), with the glycosylated Asn residues represented by green spheres. The corresponding glycans linked to Asn residues of the depicted MRH domains (chains A and B) are shortened as white sticks for clarity. (*D*) Hydrogen bonds (dashed lines) between mannose residues (green sticks) and side chains of clawed frog GNPTG in chain A (*Left*) and B (*Right*). Distances are below 3.6 Å, except for the contact with Arg97 (3.8 Å). Human residue numbers are shown throughout this figure. The types of glycosidic bonds within the glycans are indicated. (*E*) Superimposition of loops 1 and 2 of GNPTG, cation-dependent mannose-6-phosphate receptor (CDMPR, PDB ID 1C39) (51, 52), cation-independent mannose-6-phosphate receptor (CIMPR) domains 3 (PDB ID 1SYO) (53, 54), 5 (PDB ID 6P8I) (55, 56) and 9 (PDB ID 6Z30) (57), glucosidase 2 β subunit (PDB ID 4XQM) (58, 59), and lectin OS-9 (PDB ID 3AIH) (60). Two residues from GNPTG and one from glucosidase 2 β are labeled. (*F*) DMAP1 domain of GNPTG, with residues forming the hydrophobic core as white sticks and acidic side chains as red spheres.

this loop is short and bent away from the ligand, suggesting that the protein could accommodate an already-tagged mannose residue to enable second tagging.

In the MRH-containing enzyme glucosidase 2 β, a tryptophan residue near the mannose-binding site was proposed to interact with another region of that protein for proper function (58, 59). Interestingly, in GNPTG there are two entirely exposed hydrophobic side chains (Phe96 and Trp98, fully conserved in vertebrates) on a loop near the ligand-binding site (Fig. 3*E*). It is possible that this loop interacts with lysosomal hydrolases or, more likely, with the GNPT catalytic domain in order to position the glycan into the active site. The potential function of this loop remains to be determined. However, the main contact surface between the γ subunit and the GNPT complex was found to be in its N-terminal segment which precedes the MRH domain (66) and is mostly disordered in the crystal structure.

Besides the MRH domain, GNPTG contains a DMAP1-interaction module (Fig. 3*A*) (19) that shares low sequence identity with other proteins, including with the one found in GNPTAB. These domains recognize features and bind to lysosomal hydrolases (17, 19). In GNPTG, this module consists of two antiparallel α-helices interacting with each other and with

the MRH domain through a cluster of hydrophobic side chains (Fig. 3*F*). Structures of DMAP1-interaction domains from other proteins have not been determined, and the way this module recognizes its binding partners is unknown. However, positively charged lysine residues on the surface of lysosomal hydrolases play an important role in their recognition by the GNPT complex (30–35). The γ subunit is composed of 20% acidic residues (Fig. 3*F*) that form a negatively charged surface (*SI Appendix*, Fig. S3*E*) which may carry out its function.

In the GNPT $\alpha_2\beta_2\gamma_2$ complex, the γ subunits homodimerize via a disulfide bond at Cys245 (24, 67). This residue is located in a predicted coiled-coil motif that mediates dimerization. This segment is not visible in the GNPTG crystal structure (Fig. 3*A*), suggesting that it is flexible relative to the bulk of the protein. Nevertheless, the γ subunit crystallized as a dimer, with an 861-$Å^2$ interaction surface composed of contacts between the two MRH domains and between the DMAP1 modules and their opposing MRH domains (Fig. 3*C*). The two mannose-binding sites are at opposite ends of the dimer, separate by 75 Å. It is uncertain whether the crystal structure corresponds to the physiological dimer. In that case, only one γ subunit at a time would be able to position its bound glycan near one of the two GNPTAB active sites, due to steric hindrance.

Genetic alterations in GNPTAB and GNPTG can cause the lysosomal storage disorders mucolipidosis type II (fatal in early childhood) or types IIIα/β/γ (attenuated) (25). Mutations that occur within the boundaries of the minimal GNPTAB construct (Fig. 1*A*) are clustered at the core of the catalytic domain, with relatively few substitutions at its periphery, in the EF-hand domain, or at the dimerization interfaces (*SI Appendix*, Fig. S4*A*). All known missense mutations in GNPTG are located in the MRH domain but do not affect residues involved in glycan binding, suggesting a destabilizing effect on the protein's fold (*SI Appendix*, Fig. S4*B*).

The three-dimensional apo structure of a drosophila GNPTAB homolog was reported during the preparation of this paper (36). *Drosophila* contains no GNPTG gene, and its GNPTAB protein lacks several noncatalytic domains, notably the DMAP1-interaction module which in humans participates in lysosomal enzyme recognition. The physiological function of drosophila GNPT is unknown. Nevertheless, the fold of its catalytic and EF-hand domains is similar to that of zebrafish GNPT reported here (root-mean-square deviation of 2.3 Å), and the active-site residues as well as dimeric arrangement are conserved (*SI Appendix*, Fig. S4*C*). Additionally, during the review of this paper, the three-dimensional apo structure of human active GNPTAB, determined by cryogenic electron microscopy, was reported (68). Although the entire luminal portion was analyzed, only the catalytic core—approximately corresponding to the current crystal structure—was sufficiently ordered for high-resolution modeling (*SI Appendix*, Fig. S4*C*). A minor, auto-inhibited conformation was identified, in which the α subunit C-terminal portion blocks the putative catalytic pocket. A comparison of this inhibited form with the UDP-GlcNAc–bound minimal GNPT structure presented here confirms that the segment overlaps with the donor substrate position (*SI Appendix*, Fig. S4*C*). The inhibitory segment is notably excluded from the minimal construct sequence. Moreover, low-resolution density was observed for the noncatalytic domains of human GNPTAB, suggesting that these domains are flexible relative to the catalytic core, at least in the absence of a bound substrate protein.

In summary, the current study reveals the structure of the GNPT catalytic domain, its complex with UDP-GlcNAc, and a likely mechanism for the transfer reaction. The structure and mannose-binding mode of the γ subunit was also determined. These results may help assessing the impact of genetic mutations and facilitate the development of lysosomal enzyme replacement therapies (39) or other treatments for mucolipidosis (27). However, the most intriguing aspect of this complex—the basis for its ability to specifically recognize around 70 lysosomal hydrolases among the thousands of glycoproteins in the secretory pathway—remains largely unanswered. Additional structural studies of GNPT with its substrate proteins are required to uncover the details of these interactions.

Materials and Methods

Protein Constructs, Expression, and Purification. GNPTAB and GNPTG were expressed as secreted proteins in *Sf*9 insect cells infected with recombinant baculovirus. For clawed frog GNPTG (*Xenopus laevis*, UniProt ID Q68F17), the endogenous signal peptide (residues 1 to 20) was replaced by the melittin signal peptide MKFLVNVALVFMVVYISYIYA followed by a hexahistidine tag DRHHHHHHKL. The zebrafish GNPTAB minimal construct (*Danio rerio*, UniProt ID Q5RGJ8) comprised the melittin signal peptide and a hexahistidine tag DRHHHHHHGS followed by the luminal portion of the protein (residues 44 to 1173) with the segment 92 to 324 replaced by the linker PKLIKEVTELKRSKRD-PLIPECQGKQTPEKDKCYRDDNT from *Dictyostelium discoideum* (RefSeq ID XP_638036) (41) and the portion 443 to 893 deleted. This construct was based on its human homolog (39). These proteins were cloned downstream of a polyhedrin promoter. The hamster GNPTAB-GNPTG complex (*Cricetulus griseus*) consisted of the luminal portion of GNPTAB (RefSeq ID XP_027248365 with variant L806P, residues 39 to 1189, preceded by a melittin signal peptide and residues DGS, with a polyhedrin promoter) coexpressed with GNPTG (UniProt ID G3HBX0, residues 25 to 307, preceded by a melittin signal peptide and residue D, with a C-terminal hexahistidine tag HHHHHH, and a p10 promoter). To produce proteolytically processed soluble GNPTAB in insect cells, the site-1 protease cleavage site between the α and β subunits TGRQLK (15) was replaced by a furin cleavage sequence RARYKR (14), within the GNPTAB-GNPTG coexpression construct. All constructs were fully sequenced.

Expression was carried out at 27 °C for 66 h. Proteins were isolated from expression culture media by immobilized metal affinity chromatography, with 0.5 mM phenylmethylsulfonyl fluoride included in the elution buffer for the GNPTAB–GNPTG complex. Proteins were purified by size-exclusion chromatography (SEC) on Superdex 200 Increase or Superose 6 Increase columns (Cytiva) in buffer (10 mM Tris·HCl, pH 7.5) with 200 mM NaCl for GNPTG, or 100 mM NaCl and 1 mM $CaCl_2$ for GNPTAB or the two-protein complex. The complex was additionally loaded on a Mono Q anion exchange chromatography column (Cytiva) in SEC buffer and eluted with a gradient of 100 to 200 mM NaCl in SEC buffer. All proteins were concentrated to 10 mg/mL and flash-frozen.

X-Ray Crystallography. Crystallization of multiple mammalian and vertebrate GNPTAB (luminal portion) and GNPTAB–GNPTG complex homologs was attempted unsuccessfully. The GNPTAB minimal constructs from several mammalian species yielded low-resolution crystals. The zebrafish GNPTAB minimal construct crystallized at 22 °C by sitting-drop vapor diffusion in 13% polyethylene glycol (PEG) 3350 with 100 mM Hepes-Na, pH 7.5. The crystal was soaked in well solution with 10% glycerol for 60 s, then in well solution with 20% glycerol for a few seconds, and flash-frozen. X-ray diffraction data were collected at 100 K on beamline 08B1 with a Pilatus3 S 6M detector at the Canadian Macromolecular Crystallography Facility (CMCF), Canadian Light Source.

These crystals were also incubated with UDP-*N*-acetylglucosamine (UDP-GlcNAc, Sigma U4375) in soaking solution (85 mM Tris·HCl, pH 7.5, 12% PEG 3350, 50 mM UDP-GlcNAc-Na_2, and 50 mM $MgCl_2$) for 16 h at 22 °C and cryoprotected as above. X-ray diffraction data were collected at 100 K on beamline 24-ID-E with an EIGER 16M detector at the Northeastern Collaborative Access Team facility, Advanced Photon Source.

Isolated mammalian GNPTG could not be produced at high yield in insect cells. Clawed frog GNPTG crystallized in 20% PEG 3350 with 100 mM bis-Tris propane·HCl, pH 8, and 200 mM KNO_3. The crystal was soaked in well solution with 20% glycerol for a few seconds and flash-frozen. X-ray diffraction data were collected at 100 K on beamline 5.0.2 with a Pilatus3 S 6M detector at the

Berkeley Center for Structural Biology, Advanced Light Source, Lawrence Berkeley National Laboratory.

Structure Determination. Data were processed by HKL2000 (69) with autocorrections enabled. This option applies several corrective procedures including ellipsoidal truncation which reduces high-resolution data completeness depending on anisotropy. This is made apparent in the number of reflections and completeness values in "Data collection" versus "Refinement" in *SI Appendix*, Table S1. The GNPTAB structure was solved by molecular replacement using Phaser (70) in Phenix (71), with a search model derived from an AlphaFold2 prediction (72). The GNPTG structure was solved by molecular replacement with the MRH domain of glucosidase 2 β subunit (PDB ID 4XQM) (58, 59). All structures were manually built in Coot (73). Refinement was carried out by phenix.refine (74), with noncrystallographic symmetry (NCS) restraints applied for the GNPTG structure, and translation–libration–screw parameters (TLS, 10 groups) refined for the GNPTAB complex with UDP-GlcNAc. Crystallographic data collection and structure refinement statistics are presented in *SI Appendix*, Table S1. Structural images were prepared with PyMOL (PyMOL Molecular Graphics System, Version 2.4; Schrödinger, LLC) and ChimeraX (75). Interface surface areas were calculated with PISA (76). Surface electrostatic potentials were generated with PDB2PQR (77) and APBS (78). Sequence logos were generated with WebLogo (79), and alignments were performed with MUSCLE (80).

Enzymatic Activity Assays. In vitro enzymatic activity assays were carried out on the purified hamster GNPTAB-GNPTG active (furin-cleaved) complex (*SI Appendix*, Fig. S2 *C* and *D*) because its recombinant expression was better than the human homolog. The protein was incubated at 30 to 1,000 nM with 100 mM α-methyl D-mannoside (Sigma 67770), 1 mM UDP-GlcNAc (Sigma U4375) and 25 units/mL of Quick CIP (New England Biolabs M0525) in buffer (20 mM bis-Tris·HCl, pH 7, 50 mM NaCl, and 5 mM $MgCl_2$) for 1 h at 37 °C. A high methyl-mannoside concentration was used due to its low affinity as substrate (14, 21, 37). The CIP phosphatase converts the UMP product to uridine and phosphate. Five volumes of BIOMOL Green phosphate detection reagent (Enzo Life Sciences BML-AK111) were then added, followed by incubation for 10 min and absorbance measurement at 620 nm. The reaction was quantified with a phosphate standard curve. Less than 10% of the UDP-GlcNAc was converted.

GNPT activity on acid ceramidase, a lysosomal hydrolase, was assayed in vitro. The inactive (Cys143Ala mutant) acid ceramidase from the naked mole rat was purified as previously described (81), except that 20 μM kifunensine (Toronto Research Chemicals K450000) was added to the insect cell media during protein expression, in order to obtain mostly Man_9-$GlcNAc_2$ N-linked glycans. The Asn348Ser mutant was used, which contains three N-glycosylation sites. Acid ceramidase at 1 mg/mL (23 μM) was incubated with 0.25 mg/mL (0.76 μM) GNPT as above in presence of 1 mM UDP-GlcNAc in buffer (50 mM bis-Tris·HCl, pH 7, 50 mM NaCl, and 10 mM $MgCl_2$) for 1 h at 37 °C. Reducing sodium dodecyl sulfate polyacrylamide gel electrophoresis sample buffer was added, followed by incubation at 95 °C for 5 min. The reaction was analyzed on a 10% Mini-PROTEAN TGX gel (Bio-Rad Laboratories 4561036).

Data, Materials, and Software Availability. Atomic coordinates and crystallographic structure factors have been deposited in the Protein Data Bank under the accession codes 7S69 (82), 7S6N (83), and 7SJ2 (84).

ACKNOWLEDGMENTS. Part or all of the research described in this paper was performed using beamline CMCF-BM at the Canadian Light Source, a national research facility of the University of Saskatchewan, which is supported by the Canada Foundation for Innovation (CFI), the Natural Sciences and Engineering Research Council (NSERC), the National Research Council (NRC), the Canadian Institutes of Health Research (CIHR), the Government of Saskatchewan, and the University of Saskatchewan. This work is based upon research conducted at the Northeastern Collaborative Access Team beamlines, which are funded by the National Institute of General Medical Sciences from the NIH (P30 GM124165). The Eiger 16M detector on the 24-ID-E beam line is funded by an NIH-ORIP HEI grant (S10OD021527). This research used resources of the Advanced Photon Source, a US Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under contract DE-AC02-06CH11357. The Berkeley Center for Structural Biology is supported in part by the Howard Hughes Medical Institute. The Advanced Light Source is a Department of Energy Office of Science User Facility under contract DE-AC02-05CH11231. The Pilatus detector on 5.0.1. was funded under NIH grant S10OD021832. The ALS-ENABLE beamlines are supported in part by the NIH, National Institute of General Medical Sciences, grant P30 GM124169. B.N. is supported by a project grant from the CIHR (PJT-173264). A.G. is supported by a CIHR fellowship (MFE-164773).

1. UniProt Consortium, UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49** (D1), D480–D489 (2021).
2. G. Di Lorenzo *et al.*, Lysosomal proteome and secretome analysis identifies missorted enzymes and their nondegraded substrates in mucolipidosis III mouse cells. *Mol. Cell. Proteomics* **17**, 1612–1626 (2018).
3. T. Braulke, J. S. Bonifacino, Sorting of lysosomal proteins. *Biochim. Biophys. Acta* **1793**, 605–614 (2009).
4. M. F. Coutinho, M. J. Prata, S. Alves, Mannose-6-phosphate pathway: A review on its role in lysosomal function and dysfunction. *Mol. Genet. Metab.* **105**, 542–550 (2012).
5. P. Ghosh, N. M. Dahms, S. Kornfeld, Mannose 6-phosphate receptors: New twists in the tale. *Nat. Rev. Mol. Cell Biol.* **4**, 202–212 (2003).
6. R. Pohlmann, A. Waheed, A. Hasilik, K. von Figura, Synthesis of phosphorylated recognition marker in lysosomal enzymes is located in the cis part of Golgi apparatus. *J. Biol. Chem.* **257**, 5323–5325 (1982).
7. M. L. Reitman, S. Kornfeld, UDP-N-acetylglucosamine:glycoprotein N-acetylglucosamine-1-phosphotransferase. Proposed enzyme for the phosphorylation of the high mannose oligosaccharide units of lysosomal enzymes. *J. Biol. Chem.* **256**, 4275–4281 (1981).
8. M. L. Reitman, S. Kornfeld, Lysosomal enzyme targeting. N-Acetylglucosaminylphosphotransferase selectively phosphorylates native lysosomal enzymes. *J. Biol. Chem.* **256**, 11977–11980 (1981).
9. A. Varki, S. Kornfeld, Identification of a rat liver alpha-N-acetylglucosaminyl phosphodiesterase capable of removing "blocking" alpha-N-acetylglucosamine residues from phosphorylated high mannose oligosaccharides of lysosomal enzymes. *J. Biol. Chem.* **255**, 8398–8401 (1980).
10. A. Waheed, A. Hasilik, K. von Figura, Processing of the phosphorylated recognition marker in lysosomal enzymes. Characterization and partial purification of a microsomal alpha-N-acetylglucosaminyl phosphodiesterase. *J. Biol. Chem.* **256**, 5717–5721 (1981).
11. A. Gorelik, K. Illes, B. Nagar, Crystal Structure of the Mannose-6-Phosphate Uncovering Enzyme. *Structure* **28**, 426–436.e3 (2020).
12. M. Bao, J. L. Booth, B. J. Elmendorf, W. M. Canfield, Bovine UDP-N-acetylglucosamine:lysosomal-enzyme N-acetylglucosamine-1-phosphotransferase. I. Purification and subunit structure. *J. Biol. Chem.* **271**, 31437–31445 (1996).
13. M. Kudo *et al.*, The alpha- and beta-subunits of the human UDP-N-acetylglucosamine:lysosomal enzyme N-acetylglucosamine-1-phosphotransferase [corrected] are encoded by a single cDNA. *J. Biol. Chem.* **280**, 36141–36149 (2005).
14. M. Kudo, W. M. Canfield, 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 (2006).
15. K. Marschner, K. Kollmann, M. Schweizer, T. Braulke, S. Pohl, A key enzyme in the biogenesis of lysosomes is a protease that regulates cholesterol metabolism. *Science* **333**, 87–90 (2011).
16. S. Tiede *et al.*, Mucolipidosis II is caused by mutations in GNPTA encoding the alpha/beta GlcNAc-1-phosphotransferase. *Nat. Med.* **11**, 1109–1112 (2005).
17. Y. Qian, H. Flanagan-Steet, E. van Meel, R. Steet, S. A. Kornfeld, 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 (2013).
18. Y. Qian *et al.*, 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 (2015).
19. E. van Meel *et al.*, Multiple domains of GlcNAc-1-phosphotransferase mediate recognition of lysosomal enzymes. *J. Biol. Chem.* **291**, 8295–8307 (2016).
20. W. S. Lee, B. J. Payne, C. M. Gelfman, P. Vogel, S. Kornfeld, Murine UDP-GlcNAc:lysosomal enzyme N-acetylglucosamine-1-phosphotransferase lacking the gamma-subunit retains substantial activity toward acid hydrolases. *J. Biol. Chem.* **282**, 27198–27203 (2007).
21. Y. Qian *et al.*, Functions of the alpha, beta, and gamma subunits of UDP-GlcNAc:lysosomal enzyme N-acetylglucosamine-1-phosphotransferase. *J. Biol. Chem.* **285**, 3360–3370 (2010).
22. M. L. Reitman, A. Varki, S. Kornfeld, Fibroblasts from patients with I-cell disease and pseudo-Hurler polydystrophy are deficient in uridine 5′-diphosphate-N-acetylglucosamine: Glycoprotein N-acetylglucosaminylphosphotransferase activity. *J. Clin. Invest.* **67**, 1574–1579 (1981).
23. A. Hasilik, A. Waheed, K. von Figura, Enzymatic phosphorylation of lysosomal enzymes in the presence of UDP-N-acetylglucosamine. Absence of the activity in I-cell fibroblasts. *Biochem. Biophys. Res. Commun.* **98**, 761–767 (1981).
24. A. Raas-Rothschild *et al.*, Molecular basis of variant pseudo-hurler polydystrophy (mucolipidosis IIIC). *J. Clin. Invest.* **105**, 673–681 (2000).
25. R. V. Velho *et al.*, 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 (2019).
26. A. R. Ko *et al.*, AAV8-mediated expression of N-acetylglucosamine-1-phosphate transferase attenuates bone loss in a mouse model of mucolipidosis II. *Mol. Genet. Metab.* **117**, 447–455 (2016).
27. L. Matos *et al.*, Development of an antisense oligonucleotide-mediated exon skipping therapeutic strategy for mucolipidosis II: Validation at RNA level. *Hum. Gene Ther.* **31**, 775–783 (2020).
28. S. A. Khan, S. C. Tomatsu, Mucolipidoses overview: Past, present, and future. *Int. J. Mol. Sci.* **21**, 6812 (2020).
29. A. Waheed, A. Hasilik, K. von Figura, UDP-N-acetylglucosamine:lysosomal enzyme precursor N-acetylglucosamine-1-phosphotransferase. Partial purification and characterization of the rat liver Golgi enzyme. *J. Biol. Chem.* **257**, 12322–12331 (1982).

30. T. J. Baranski, P. L. Faust, S. Kornfeld, Generation of a lysosomal enzyme targeting signal in the secretory protein pepsinogen. *Cell* **63**, 281–291 (1990).
31. J. W. Cuozzo, K. Tao, Q. L. Wu, W. Young, G. G. Sahagian, Lysine-based structure in the proregion of procathepsin L is the recognition site for mannose phosphorylation. *J. Biol. Chem.* **270**, 15611–15619 (1995).
32. J. W. Cuozzo, G. G. Sahagian, Lysine is a common determinant for mannose phosphorylation of lysosomal proteins. *J. Biol. Chem.* **269**, 14490–14496 (1994).
33. R. Tikkanen, M. Peltola, C. Oinonen, J. Rouvinen, L. Peltonen, Several cooperating binding sites mediate the interaction of a lysosomal enzyme with phosphotransferase. *EMBO J.* **16**, 6684–6693 (1997).
34. J. W. Cuozzo, K. Tao, M. Cygler, J. S. Mort, G. G. Sahagian, Lysine-based structure responsible for selective mannose phosphorylation of cathepsin D and cathepsin L defines a common structural motif for lysosomal enzyme targeting. *J. Biol. Chem.* **273**, 21067–21076 (1998).
35. R. Steet, W. S. Lee, S. Kornfeld, Identification of the minimal lysosomal enzyme recognition domain in cathepsin D. *J. Biol. Chem.* **280**, 33318–33323 (2005).
36. S. Du *et al.*, Structural insights into how GlcNAc-1-phosphotransferase directs lysosomal protein transport. *J. Biol. Chem.* **298**, 101702 (2022).
37. M. Bao, B. J. Elmendorf, J. L. Booth, R. R. Drake, W. M. Canfield, Bovine UDP-N-acetylglucosamine:lysosomal-enzyme N-acetylglucosamine-1-phosphotransferase. II. Enzymatic characterization and identification of the catalytic subunit. *J. Biol. Chem.* **271**, 31446–31451 (1996).
38. P. Sperisen, C. D. Schmid, P. Bucher, O. Zilian, Stealth proteins: In silico identification of a novel protein family rendering bacterial pathogens invisible to host immune defense. *PLOS Comput. Biol.* **1**, e63 (2005).
39. L. Liu, W. S. Lee, B. Doray, S. Kornfeld, Engineering of GlcNAc-1-phosphotransferase for production of highly phosphorylated lysosomal enzymes for enzyme replacement therapy. *Mol. Ther. Methods Clin. Dev.* **5**, 59–65 (2017).
40. C. Maris, C. Dominguez, F. H. Allain, The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J.* **272**, 2118–2131 (2005).
41. L. Liu, W. S. Lee, B. Doray, S. Kornfeld, Role of spacer-1 in the maturation and function of GlcNAc-1-phosphotransferase. *FEBS Lett.* **591**, 47–55 (2017).
42. W. R. Gordon, K. L. Arnett, S. C. Blacklow, The molecular logic of Notch signaling–a structural and biochemical perspective. *J. Cell Sci.* **121**, 3109–3119 (2008).
43. A. F. Williams, A. N. Barclay, The immunoglobulin superfamily–domains for cell surface recognition. *Annu. Rev. Immunol.* **6**, 381–405 (1988).
44. R. De Pace *et al.*, Subunit interactions of the disease-related hexameric GlcNAc-1-phosphotransferase complex. *Hum. Mol. Genet.* **24**, 6826–6835 (2015).
45. A. Lewit-Bentley, S. Réty, EF-hand calcium-binding proteins. *Curr. Opin. Struct. Biol.* **10**, 637–643 (2000).
46. T. Danyukova *et al.*, 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 (2020).
47. L. Zheng *et al.*, Cryo-EM structures of human GMPPA-GMPPB complex reveal how cells maintain GDP-mannose homeostasis. *Nat. Struct. Mol. Biol.* **28**, 1–12 (2021).
48. U. M. Unligil *et al.*, X-ray crystal structure of rabbit N-acetylglucosaminyltransferase I: Catalytic mechanism and a new protein superfamily. *EMBO J.* **19**, 5269–5280 (2000).
49. S. Munro, The MRH domain suggests a shared ancestry for the mannose 6-phosphate receptors and other N-glycan-recognising proteins. *Curr. Biol.* **11**, R499–R501 (2001).
50. C. D'Alessio, N. M. Dahms, Glucosidase II and MRH-domain containing proteins in the secretory pathway. *Curr. Protein Pept. Sci.* **16**, 31–48 (2015).
51. D. L. Roberts, D. J. Weix, N. M. Dahms, J. J. Kim, Molecular basis of lysosomal enzyme recognition: Three-dimensional structure of the cation-dependent mannose 6-phosphate receptor. *Cell* **93**, 639–648 (1998).
52. L. J. Olson, J. Zhang, Y. C. Lee, N. M. Dahms, J. J. Kim, Structural basis for recognition of phosphorylated high mannose oligosaccharides by the cation-dependent mannose 6-phosphate receptor. *J. Biol. Chem.* **274**, 29889–29896 (1999).
53. L. J. Olson, R. D. Yammani, N. M. Dahms, J. J. Kim, Structure of uPAR, plasminogen, and sugar-binding sites of the 300 kDa mannose 6-phosphate receptor. *EMBO J.* **23**, 2019–2028 (2004).
54. L. J. Olson, N. M. Dahms, J. J. Kim, The N-terminal carbohydrate recognition site of the cation-independent mannose 6-phosphate receptor. *J. Biol. Chem.* **279**, 34000–34009 (2004).
55. L. J. Olson *et al.*, Structural basis for recognition of phosphodiester-containing lysosomal enzymes by the cation-independent mannose 6-phosphate receptor. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 12493–12498 (2010).
56. L. J. Olson *et al.*, Allosteric regulation of lysosomal enzyme recognition by the cation-independent mannose 6-phosphate receptor. *Commun. Biol.* **3**, 498 (2020).
57. A. J. Bochel *et al.*, Structure of the human cation-independent mannose 6-phosphate/IGF2 receptor domains 7-11 uncovers the mannose 6-phosphate binding site of domain 9. *Structure* **28**, 1300–1312.e5 (2020).
58. L. J. Olson *et al.*, Structure of the lectin mannose 6-phosphate receptor homology (MRH) domain of glucosidase II, an enzyme that regulates glycoprotein folding quality control in the endoplasmic reticulum. *J. Biol. Chem.* **288**, 16460–16475 (2013).
59. L. J. Olson *et al.*, Crystal structure and functional analyses of the lectin domain of glucosidase II: Insights into oligomannose recognition. *Biochemistry* **54**, 4097–4111 (2015).
60. T. Satoh *et al.*, Structural basis for oligosaccharide recognition of misfolded glycoproteins by OS-9 in ER-associated degradation. *Mol. Cell* **40**, 905–916 (2010).
61. A. Varki, S. Kornfeld, Structural studies of phosphorylated high mannose-type oligosaccharides. *J. Biol. Chem.* **255**, 10847–10858 (1980).
62. R. N. Bohnsack *et al.*, Cation-independent mannose 6-phosphate receptor: A composite of distinct phosphomannosyl binding sites. *J. Biol. Chem.* **284**, 35215–35226 (2009).
63. L. O. Tremblay, A. Herscovics, Cloning and expression of a specific human alpha 1,2-mannosidase that trims Man9GlcNAc2 to Man8GlcNAc2 isomer B during N-glycan biosynthesis. *Glycobiology* **9**, 1073–1078 (1999).
64. C. Reily, T. J. Stewart, M. B. Renfrow, J. Novak, Glycosylation in health and disease. *Nat. Rev. Nephrol.* **15**, 346–366 (2019).
65. D. E. Sleat, H. Zheng, M. Qian, P. Lobel, Identification of sites of mannose 6-phosphorylation on lysosomal proteins. *Mol. Cell. Proteomics* **5**, 686–701 (2006).
66. R. V. Velho, R. De Pace, H. Tidow, T. Braulke, S. Pohl, Identification of the interaction domains between α- and γ-subunits of GlcNAc-1-phosphotransferase. *FEBS Lett.* **590**, 4287–4295 (2016).
67. M. Encarnação, K. Kollmann, M. Trusch, T. Braulke, S. Pohl, Post-translational modifications of the gamma-subunit affect intracellular trafficking and complex assembly of GlcNAc-1-phosphotransferase. *J. Biol. Chem.* **286**, 5311–5318 (2011).
68. H. Li *et al.*, Structure of the human GlcNAc-1-phosphotransferase αβ subunits reveals regulatory mechanism for lysosomal enzyme glycan phosphorylation. *Nat. Struct. Mol. Biol.* **29**, 348–356 (2022).
69. Z. Otwinowski, W. Minor, Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
70. A. J. McCoy *et al.*, Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
71. D. Liebschner *et al.*, Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallogr. D Struct. Biol.* **75**, 861–877 (2019).
72. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
73. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
74. P. V. Afonine *et al.*, Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367 (2012).
75. E. F. Pettersen *et al.*, UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).
76. E. Krissinel, K. Henrick, Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
77. T. J. Dolinsky, J. E. Nielsen, J. A. McCammon, N. A. Baker, PDB2PQR: An automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* **32**, W665–W667 (2004).
78. N. A. Baker, D. Sept, S. Joseph, M. J. Holst, J. A. McCammon, Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10037–10041 (2001).
79. G. E. Crooks, G. Hon, J. M. Chandonia, S. E. Brenner, WebLogo: A sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
80. R. C. Edgar, MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
81. A. Gebai, A. Gorelik, Z. Li, K. Illes, B. Nagar, Structural basis for the activation of acid ceramidase. *Nat. Commun.* **9**, 1621 (2018).
82. A. Gorelik, K. Illes, K. H. Bui, B. Nagar, N-acetylglucosamine-1-phosphotransferase (GNPT) gamma subunit (GNPTG), from clawed frog. Protein Data Bank. https://www.rcsb.org/structure/7S69. Deposited 13 September 2021.
83. A. Gorelik, K. Illes, K. H. Bui, B. Nagar, N-acetylglucosamine-1-phosphotransferase (GNPT) alpha and beta subunits (GNPTAB) catalytic domain, from zebrafish. Protein Data Bank. https://www.rcsb.org/structure/7S6N. Deposited 14 September 2021.
84. A. Gorelik, K. Illes, K. H. Bui, B. Nagar, N-acetylglucosamine-1-phosphotransferase (GNPT) alpha and beta subunits (GNPTAB) catalytic domain, from zebrafish, in complex with uridine diphosphate N-acetylglucosamine (UDP-GlcNAc) and magnesium. Protein Data Bank. https://www.rcsb.org/structure/7SJ2. Deposited 15 October 2021.