
Figures and figure supplements

Structure of the human heparan- α -glucosaminide *N*-acetyltransferase (HGSNAT)

Vikas Navratna et al.

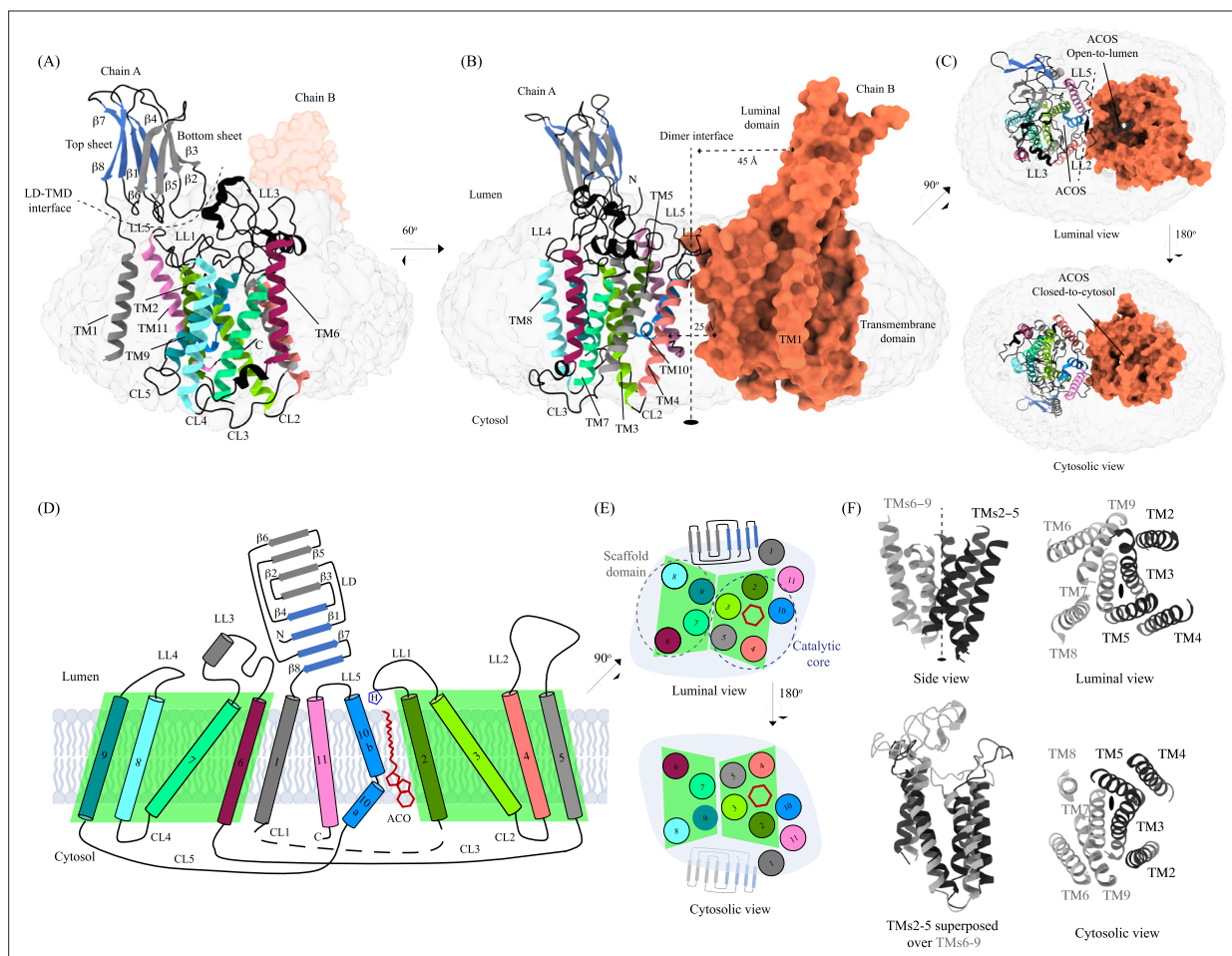


Figure 1. Structure of heparan- α -glucosaminide N-acetyltransferase (HGSNAT). Panels (A) and (B) show two different orientations of the HGSNAT dimer that highlight (dashed lines) the LD-TMD interface and dimer interface, respectively. Micelle is displayed in gray. Chain A is displayed as a cartoon and chain B as an orange surface. All the luminal loops (LLs), cytosolic loops (CLs), and the loops that connect β -sheets are shown in black. The top and bottom sheets in the luminal domain (LD) are colored blue and gray, respectively. The twofold rotation axis is displayed as a dashed line with an ellipsoid. (C) Luminal (top) and cytosolic (bottom) views of the protein. The surface representation of chain B suggests that the acetyl-CoA binding site (ACOS) is more accessible from the luminal side (top) than the cytosolic side (bottom). (D) 2D topology of HGSNAT and YeiB family. The helices and strands in the topology are colored similarly to the 3D structure. Transmembrane helices (TMs) 2–5 and 6–9 form two bundles (4+4), highlighted by green parallelograms, that are related to each other by a twofold rotation parallel to the plane of the membrane. TMs 1, 10, and 11 do not seem involved in this internal symmetry, with TM10 being bent in the plane of the membrane into two halves TM10a and TM10b. The relative position of bound ACO and active site H269 of LL1 are indicated. (E) Luminal (top) and cytosolic (bottom) views of the protein topology. TMs 2–5 and TM10 enclose ACOS (red hexagon) and are referred to as catalytic core (blue dashed oval). TMs 6–9 will be referred to as scaffold domain (gray dashed oval). (F) 4+4 bundle formed by TMs 2–5 (black) and TMs 6–9 (gray) are related by a twofold rotation. The last sub-panel (bottom left) shows a superposition of TMs 2–5 on TMs 6–9.

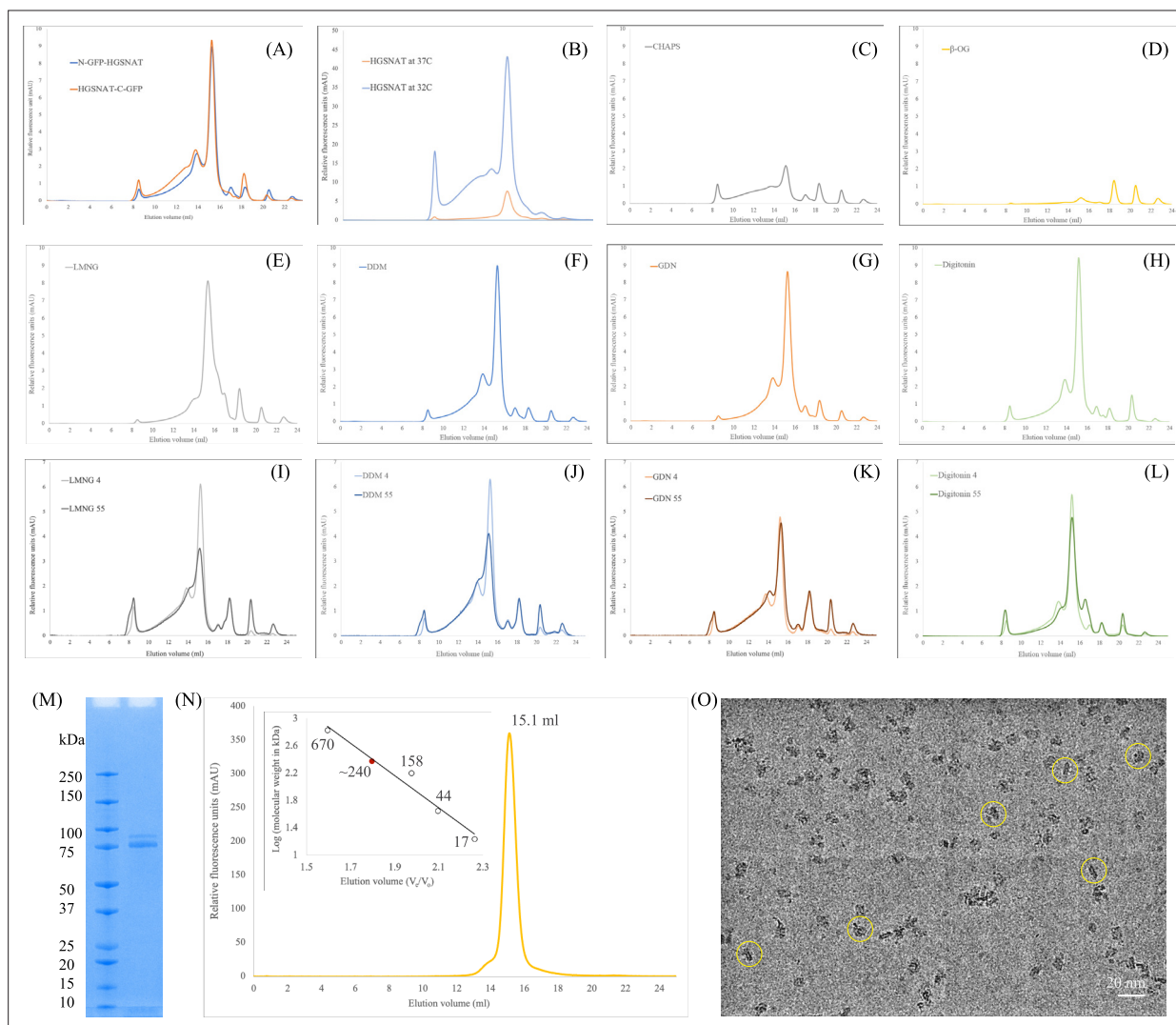


Figure 1—figure supplement 1. Purification of heparan- α -glucosaminide N-acetyltransferase (HGSNAT). **(A)** Comparison of expression of N- and C-terminal GFP fusions of Heparan- α -glucosaminide N-acetyltransferase (HGSNAT) in HEK293S GnT1⁻ cell lysates, solubilized in 1% DDM. **(B)** Comparison of relative overexpression of N-GFP-HGSNAT in cultures grown at 37°C and 32°C, post-transduction. **(C–H)** Relative solubility and homogeneity comparison in 1% of CHAPS, β -OG, LMNG, DDM, GDN, and digitonin, respectively, prepared in 25 mM Tris-HCl, pH 7.5, 200 mM NaCl, 1 mM PMSF, 0.8 μ M aprotinin, 2 μ g/mL leupeptin, and 2 μ M pepstatin A. **(I–L)** Comparison of relative thermal stability of detergent solubilized HGSNAT in 1% of LMNG, DDM, GDN, and digitonin, respectively. Samples analyzed after heat treatment at 55 °C for 15 min have been marked with a suffix 55, and samples stored in cold room are marked with a suffix 4. **(M)** SDS-PAGE (12%) showing purity and monomeric molecular weight of HGSNAT. Although, monomeric molecular weight is ~100 kDa, the full-length GFP fusion of HGSNAT, like most eukaryotic membrane proteins, displays anomalous electrophoretic mobility and runs around 75 kDa. **(N)** Intrinsic tryptophan fluorescence size-exclusion chromatogram of purified HGSNAT analyzed on Superose 6 Increase 10/300 GL column at 0.5 ml/min flowrate in LMNG-based FSEC running buffer. Red dot on the standard plot (log of protein molecular weight kDa) vs. ratio of the elution volume to the void volume (V_e/V_0) indicates that recombinant N-GFP-HGSNAT elutes at 15.1 ml corresponding to a dimer of ~240 kDa. **(O)** Representative micrograph imaged on Titan Krios using UltrAuFoil holey-gold 300 mesh 1.2/1.3 μ m grid of vitrified N-GFP-HGSNAT at 0.9 mg/ml. The protein (yellow circles) distribution on grids, along with SDS-PAGE and size-exclusion chromatogram shows a monodisperse sample preparation.

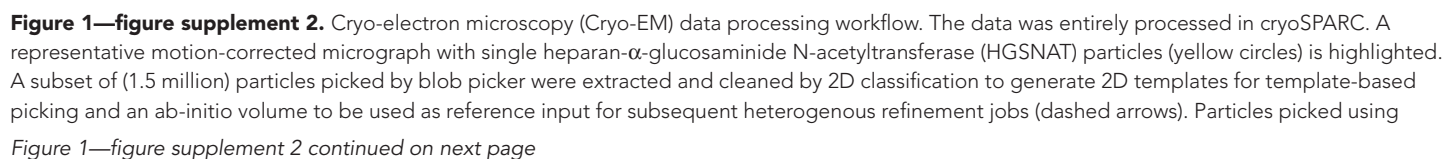


Figure 1—figure supplement 2 continued

template picker and blob picker were individually cleaned by 2D classification to remove obvious junk particles and then were pooled and duplicates were removed for sorting by heterogenous refinement, and iterative 2D and 3D classification. Throughout the processing workflow classes with most well-resolved luminal domain was used as input references for the subsequent steps of processing (highlighted by a dashed boxes). A resultant stack of 85,500 particles (representative 2D classes highlighted) was further cleaned up based on CTF fit ($<4 \text{ \AA}$) to end up with a final particle stack of about 57000 particles. C2 symmetry was applied at this stage and non-uniform and contrast transfer function (CTF) refinements were performed to yield a C2 map at 3.26 \AA . This map was used for model building and analyzing the structure of HGSNAT. A C1 map was generated by symmetry expansion of the final particle stack followed by local refinement to compare the quality of data with and without C2 symmetry application. Local refinements with masks focused on luminal domain (LD) and transmembrane domain (TMD) domain were performed separately, and a composite map was generated by combining these local refined maps to improve the density in these regions. The composite map was only used to finalize the fit of LBD side chains in the final model.

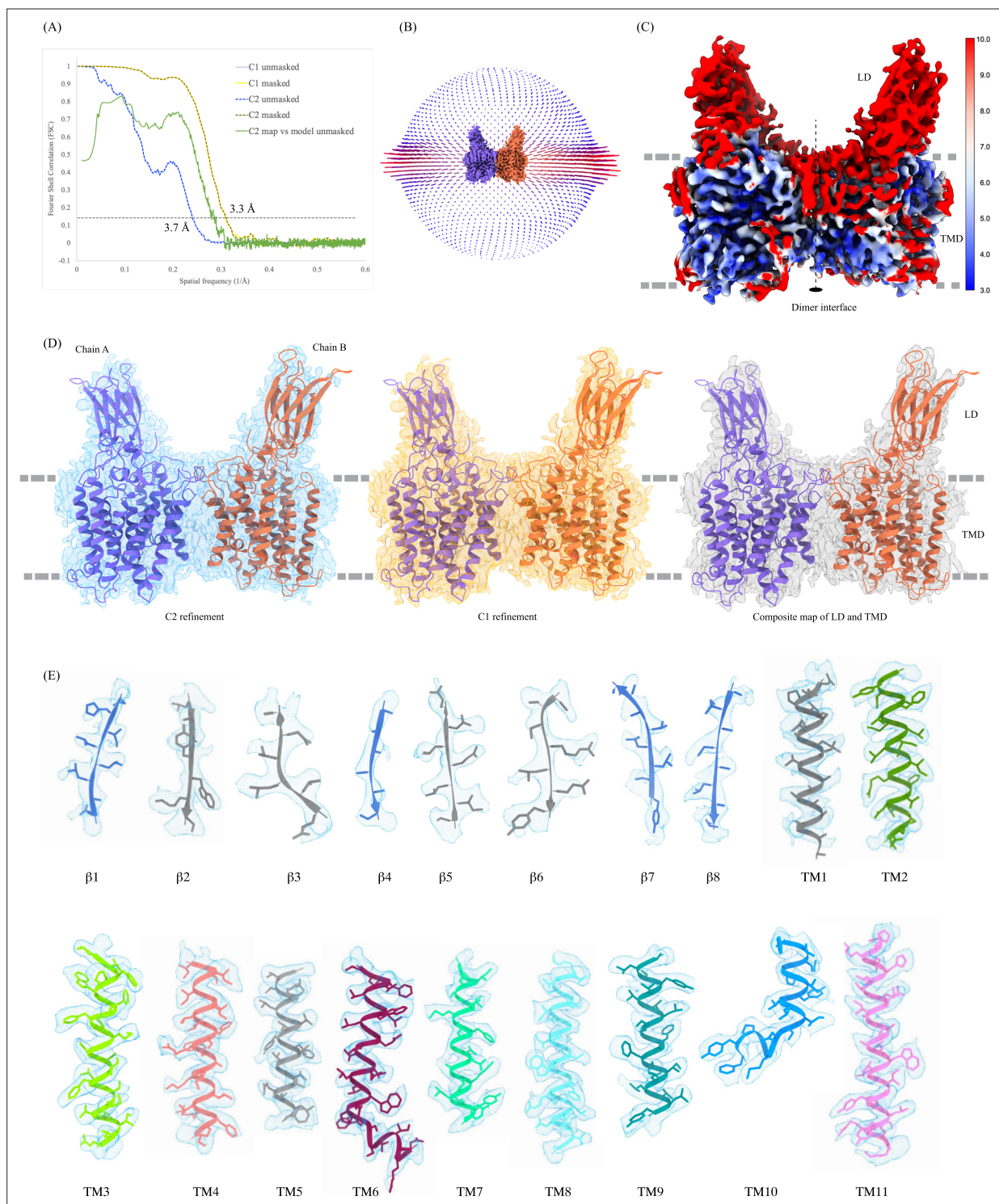


Figure 1—figure supplement 3. Cryo-electron microscopy (Cryo-EM) data quality, reconstruction, and model building. **(A)** Fourier shell coefficient (FSC) curves for cross-validation. The final masked heparan- α -glucosaminide N-acetyltransferase (HGSNAT) (C1: light yellow; C2: dark yellow dashed) and unmasked (C1: light blue; C2: dark blue dashed) refinement maps. Model vs. final C2 map unmasked (green). Gray and black dashed lines indicate FSC = 0.143 and FSC = 0.5 thresholds, respectively. FSC curves were calculated using Mtriage in Phenix. **(B)** Angular distribution of particles used in the final reconstruction. **(C)** C2 map colored by estimated local resolution. **(D)** HGSNAT modeled by ModelAngelo into the C2 map (blue). The fit of the same model in C1 (orange) and composite map (gray) of LD and TMD created in ChimeraX. All maps are displayed at level 0.21 in ChimeraX. **(E)** Cryo-EM density of all the secondary structure elements, β 1– β 8 and TMs 1–11, shown in light blue (display level between 0.18–0.25 in ChimeraX). Side chains for almost all the elements could be modeled unambiguously into the density. At places with missing density, the side chains were trimmed to C β .

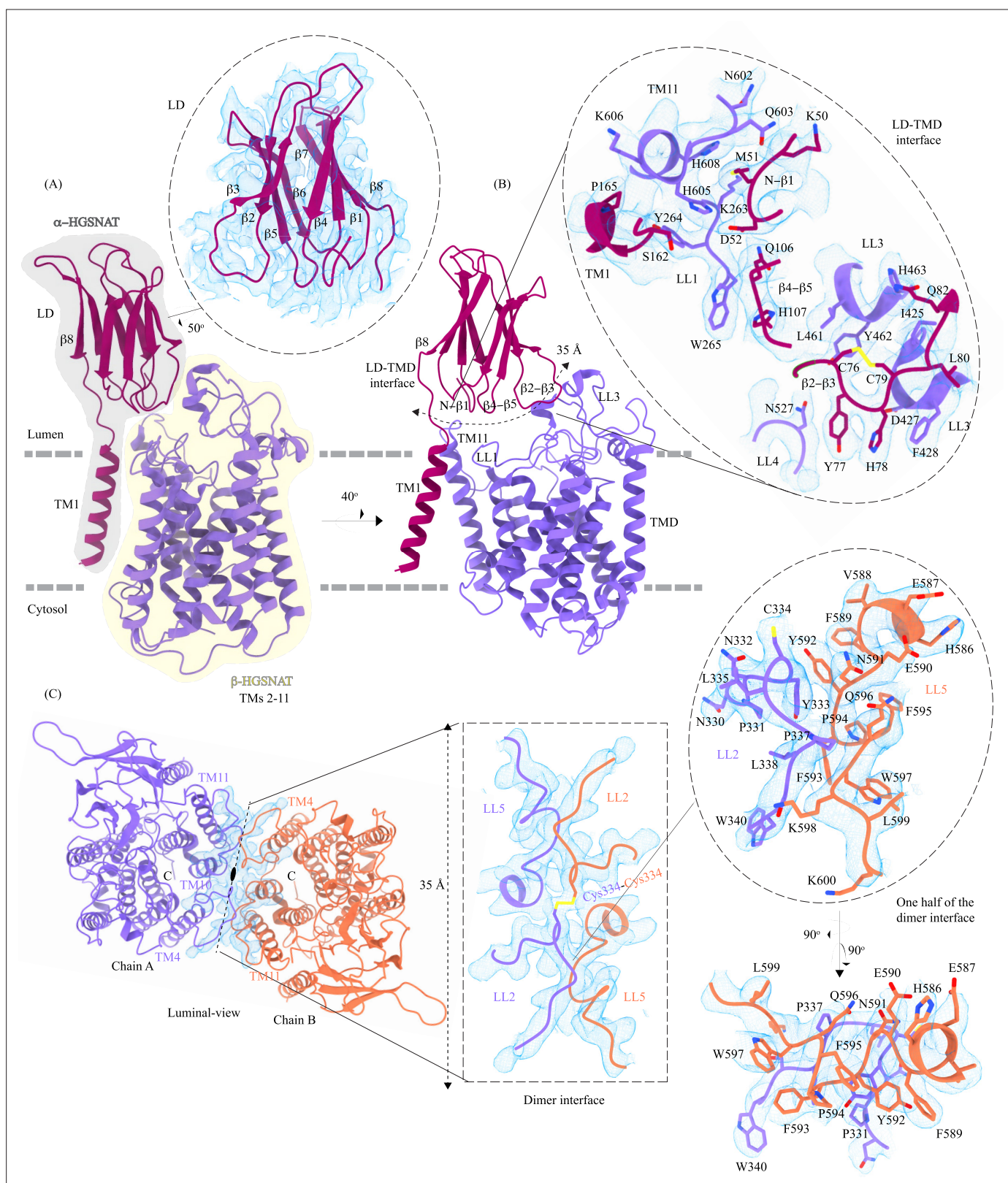


Figure 2. Domain organization, and LD-TMD and dimer interfaces of heparan- α -glucosaminide N-acetyltransferase (HGSNAT). (A) HGSNAT is predicted to be proteolyzed into two chains of unequal size - α -HGSNAT (dark magenta cartoon, gray shaded area) and β -HGSNAT (purple cartoon, yellow shaded area). The site for proteolysis remains debated. Based on our structure and prediction of HGSNAT structures from other kingdoms (Figure 2—figure supplement 1), we have represented α - and β -HGSNAT fragments as shown in panel A. The inset (dashed oval) shows the luminal view of the α -HGSNAT domain. Figure 2 continued on next page

Figure 2 continued

domain (dark magenta) fit to cryo-electron microscopy (cryo-EM) density (blue; display level 0.21 of the composite map in ChimeraX) (**Figure 1—figure supplement 3**). The lysosomal membrane is shown as a dashed gray line. **(B)** LD-TMD interface is highlighted (dashed line). Inset highlights the residues that interact at the LD-TMD interface, and cryo-EM density for the same (blue; display level 0.25 of the 3.26 Å C2 refined map in ChimeraX). C76-C79 disulfide of $\beta 2$ - $\beta 3$ turns is shown as yellow sticks, while the residue sidechains are colored the same as their secondary structure elements, with heteroatoms highlighted. **(C)** Luminal view of the protein with dimer interface highlighted (dashed line). Inset (dashed rectangle) highlights LL2 and LL5 that line the dimer interface, and the C334-C334 inter-chain disulfide (yellow) between the chains A (purple) and B (orange). The dashed oval inset shows one-half of the dimer interface with LL2 and LL5 of chains A and B, respectively, contributing other hydrophobic interactions that stabilize the dimer interface. The cryo-EM density in panel C is displayed as blue mesh (display level 0.22 of the C2 refine map in ChimeraX).

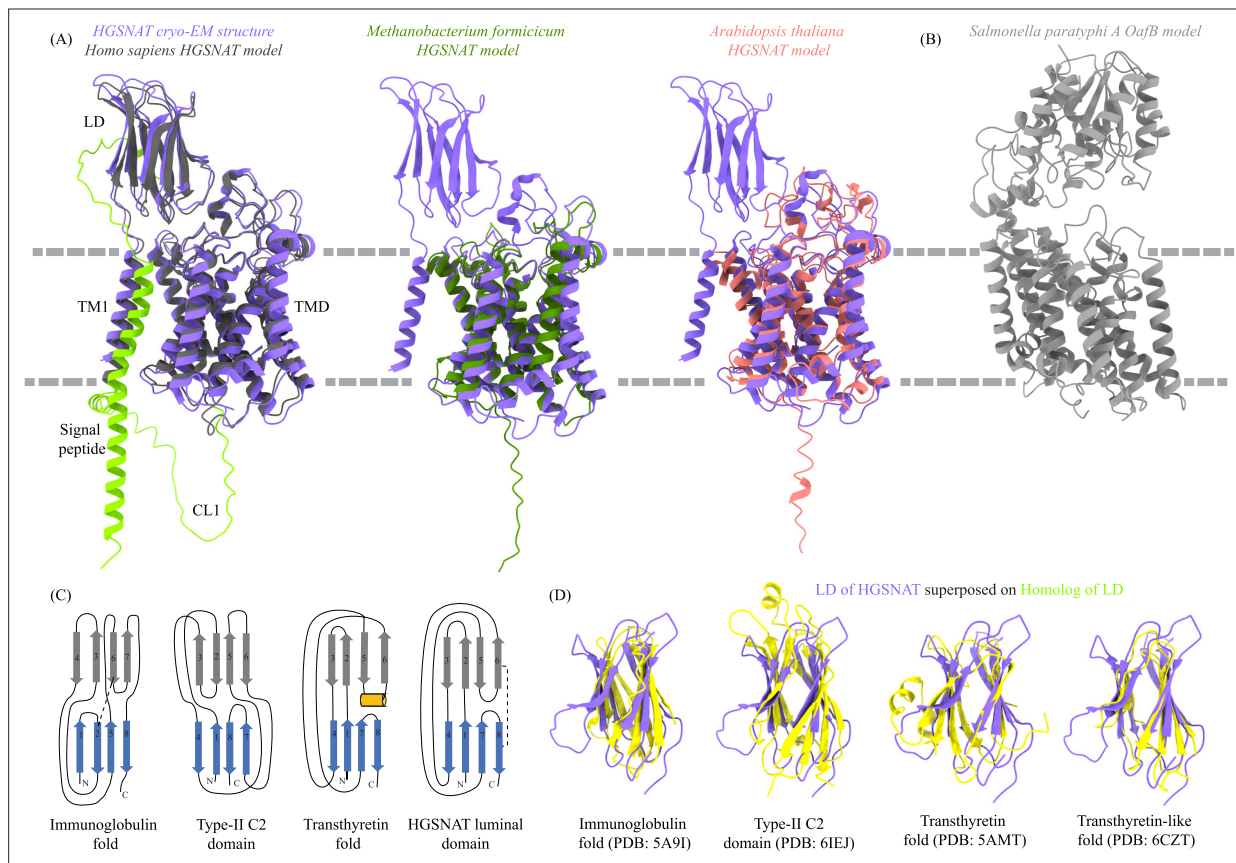


Figure 2—figure supplement 1. Homologs of heparan- α -glucosaminide N-acetyltransferase (HGSNAT). **(A)** Superposition of HGSNAT cryo-electron microscopy (cryo-EM) structure (purple) with the AlphaFold models of human (Uniprot: Q68CP4, dark gray), *Methanobacterium formicum* (Uniprot: K2QAW2, green), and *Arabidopsis thaliana* (Uniprot: A0A5S9Y8V3, pink) HGSNATs. α RMSDs of the superpositions are 1.34 Å, 1.17 Å, and 1.13 Å respectively, suggesting a conserved HGSNAT fold across different kingdoms. AlphaFold model of HGSNAT shown here is of isoform 1, that has extra 28 residues on the N-term as compared to isoform 2. The structure is of isoform 2. The cryo-EM density did not allow modeling of residues upstream of β 1 on the N-terminus and CL1, which have been highlighted yellow in the AlphaFold model. **(B)** AlphaFold model of acetyltransferase model of *Salmonella paratyphi A* OafB, an O-antigen modifying transmembrane acetyltransferase of the acyltransferase-3/acetyl-CoA transporter (ATAT) family within the TmAT superfamily (Uniprot: A0A0H2WM30). Despite predicted to be in the same superfamily as HGSNAT, a meaningful alignment and similarity to HGSNAT was not observed, highlighting the diversity of membrane bound acetyltransferases. **(C)** Comparison of topologies of immunoglobulin (Ig) fold, type-II C2 domain, and transthyretin fold with luminal domain (LD) of HGSNAT. Strands in two sheets are colored blue and gray and conserved helical turn in transthyretin fold is shown in orange. Conserved disulfides are shown as dashed lines. **(D)** Superposition of structures of Ig fold (PDB: 5A9I), type-II C2 domain (PDB: 6IEJ), transthyretin fold (PDB: 5AMT), and transthyretin-like domain (PDB: 6CZT) onto LD of HGSNAT. Based on the topology and structure superposition, it appears that LD is transthyretin-like domain.

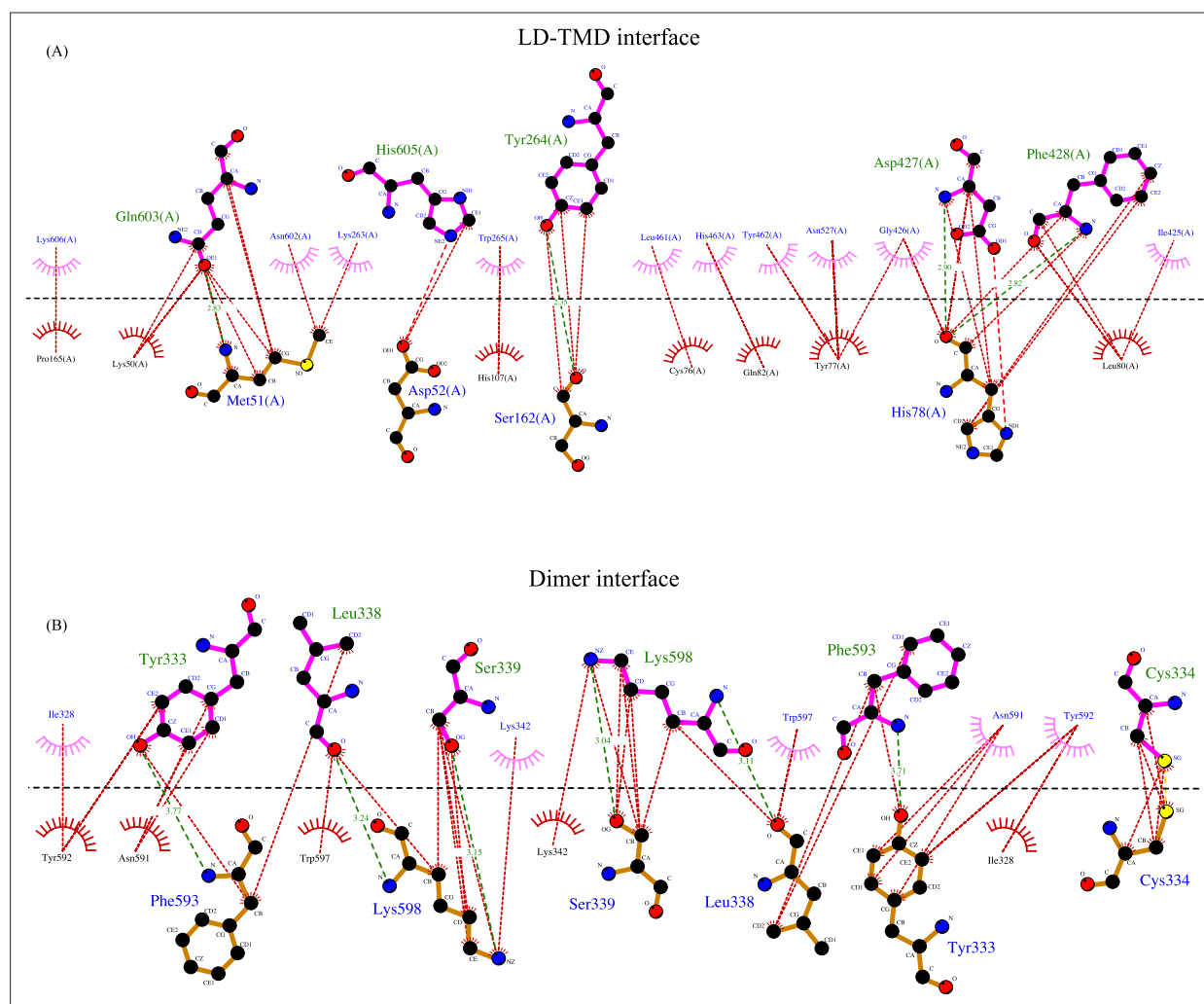


Figure 2—figure supplement 2. Interactions at LD-TMD and dimer interface. A 2D depiction of network of interactions (<4.5 Å) between residues at the LD-TMD interface (top, **A**) and at the dimer interface (bottom, **B**) generated in LigPlot+. Hydrogen bonds are shown as dashed lines with bond distance. Nonbonded and hydrophobic interactions are shown as dotted lines. Residues involved in nonbonded interactions are displayed as eye lashes. Chain A and chain B residues are shown in blue and orange. Dashed black line indicates the interface. The sulfurs of involved in disulfide bond at the dimer interface are highlighted in yellow.

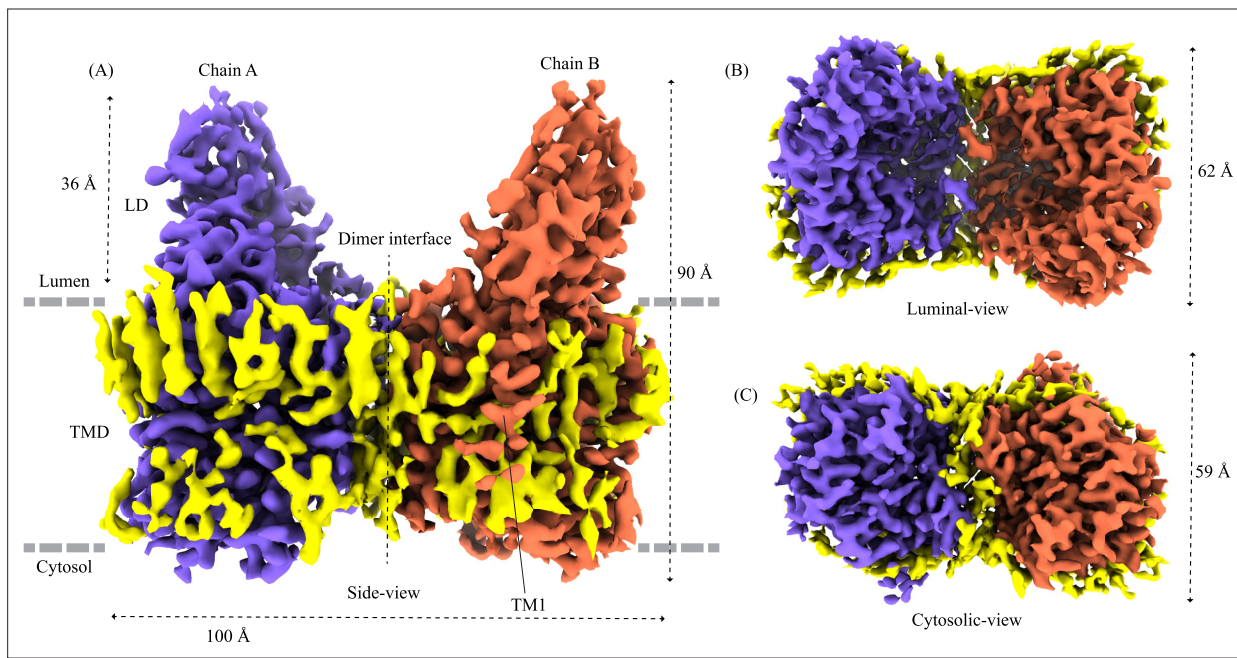


Figure 2—figure supplement 3. Lipids and detergent in the structure. Ordered density observed in our final cryo-electron microscopy (cryo-EM) map that did not account for protein and ligand has been displayed as yellow density (display level 0.22 of the C2 refine map in ChimeraX) in side-view (A), luminal-view (B), and cytosolic-view (C). We believe these are ordered lipids and detergent molecules that interact with hydrophobic patches of the protein. Towards the cytosolic side (C) we find lipid/detergent density between the two protomers, forming a partition between two acetyl-CoA binding sites (ACOSs). Chain A and chain B are shown in purple and orange. Dashed line indicates dimer interface.

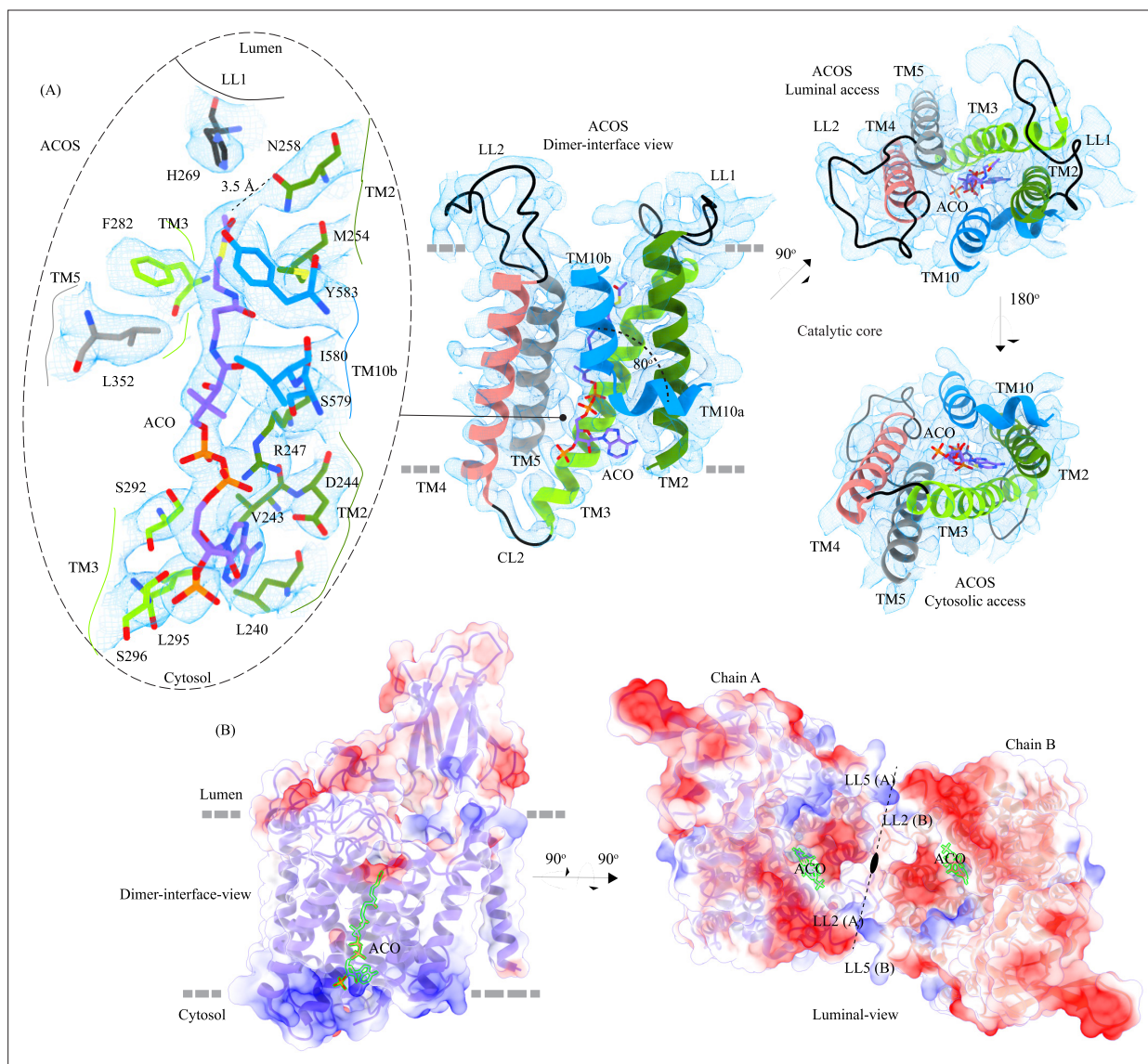


Figure 3. Acetyl-CoA binding site (ACOS). **(A)** Catalytic core (chain A) of heparan- α -glucosaminide N-acetyltransferase (HGSNAT) comprised of transmembrane helices (TMs) 2–5 and TM 10. Luminal loops (LLs) and cytosolic loops (CLs) are shown in black, and the helices are colored as in **Figure 1**. Acetyl-CoA (ACO) is colored (purple), the same as chain A in **Figure 2** with heteroatoms highlighted. The inset (dashed oval) shows acetyl-CoA binding site (ACOS) and highlights the amino acids of HGSNAT that interact with ACO. The amino acids are colored the same as the corresponding TMs, with heteroatoms highlighted. Cryo-electron microscopy (Cryo-EM) density for ACOS is displayed as blue mesh (display level 0.3 of the 3.26 Å C2 refine map in ChimeraX). ACO could be modeled to the densities at chain A and B ACOSs with a mean correlation coefficient (CC) of 0.77. The nucleoside headgroup of ACO plugs in the cytosolic access of ACOS, and the luminal access seems relatively more accessible. **(B)** Electrostatic potential and surface charge distribution of HGSNAT, with the surface display colored based on the potential contoured from -10 kT (red) to +10 kT (blue). ACO bound at the ACOS is highlighted in golden yellow. Luminal and cytosolic sides of the protein show a conspicuous polarity. The lysosomal membrane is shown as a dashed gray line in both sub-panels.

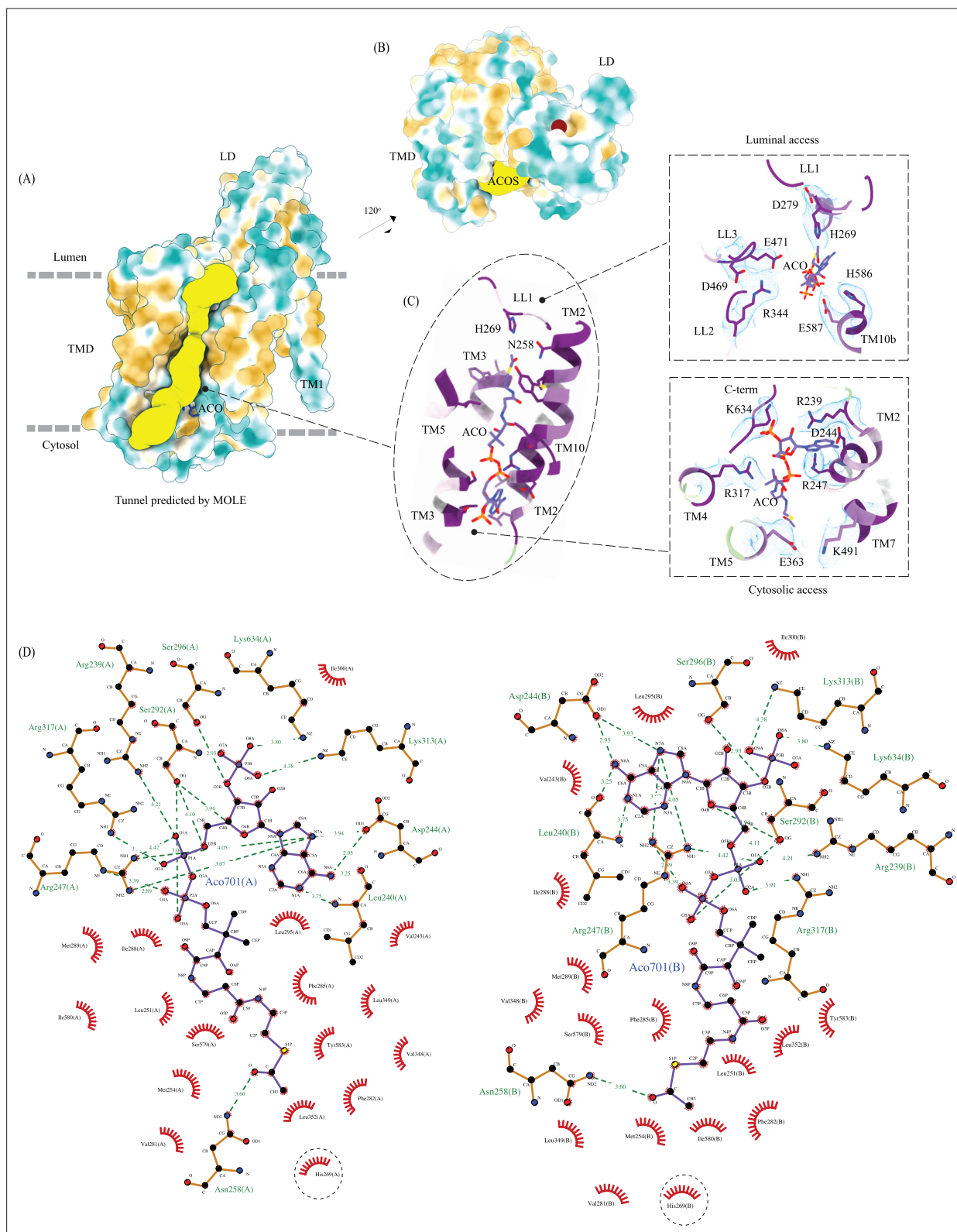


Figure 3—figure supplement 1. Ligand binding sites of heparan- α -glucosaminide N-acetyltransferase (HGSNAT). (A) Surface representation of HGSNAT (chain A), with hydrophobic and hydrophilic amino acids colored in orange and cyan, respectively. Predicted acetyl-CoA access tunnel (yellow) by MOLEonline with a probe radius of ~1.5 Å (Pravda et al., 2018). Acetyl-CoA (ACO) bound at HGSNAT is shown in blue. It is apparent that the nucleoside head group and the acetyl group interact with hydrophilic residues and the pantothenate group is supported by hydrophobic residues.

Figure 3—figure supplement 1 continued on next page

Figure 3—figure supplement 1 continued

(B) Ligand binding site on luminal domain (LD) (maroon sphere) predicted by DeepSite (**Jiménez et al., 2017**) **(C)** acetyl-CoA binding site (ACOS) color coded based on the evolutionary sequence conservation scores obtained from ConSurf server. In the insets are the integral salt-bridges of the luminal (top) and cytosolic (bottom) access of ACOS. The cryo-EM density for the salt-bridges is shown in blue (display level 0.22 of the C2 refine map in ChimeraX). **(D)** 2D depiction of the network of interactions of ACO modeled at chain A (left) and chain B (right) with HGSNAT residues that lie <4.5 Å away from ACO, generated in LigPlot+. Hydrogen bonds are depicted by dashed lines, and residues that are involved in hydrogen bonds with ACO are shown as ball & stick models. Non-bonded contacts are indicated as eye lashes. The predicted active site H269 is highlighted by dashed circle. In our structure N258 forms weak hydrogen bonds with the acetyl group of ACO. We believe that N258 holds onto ACO until H269 is protonated and ready for catalysis.

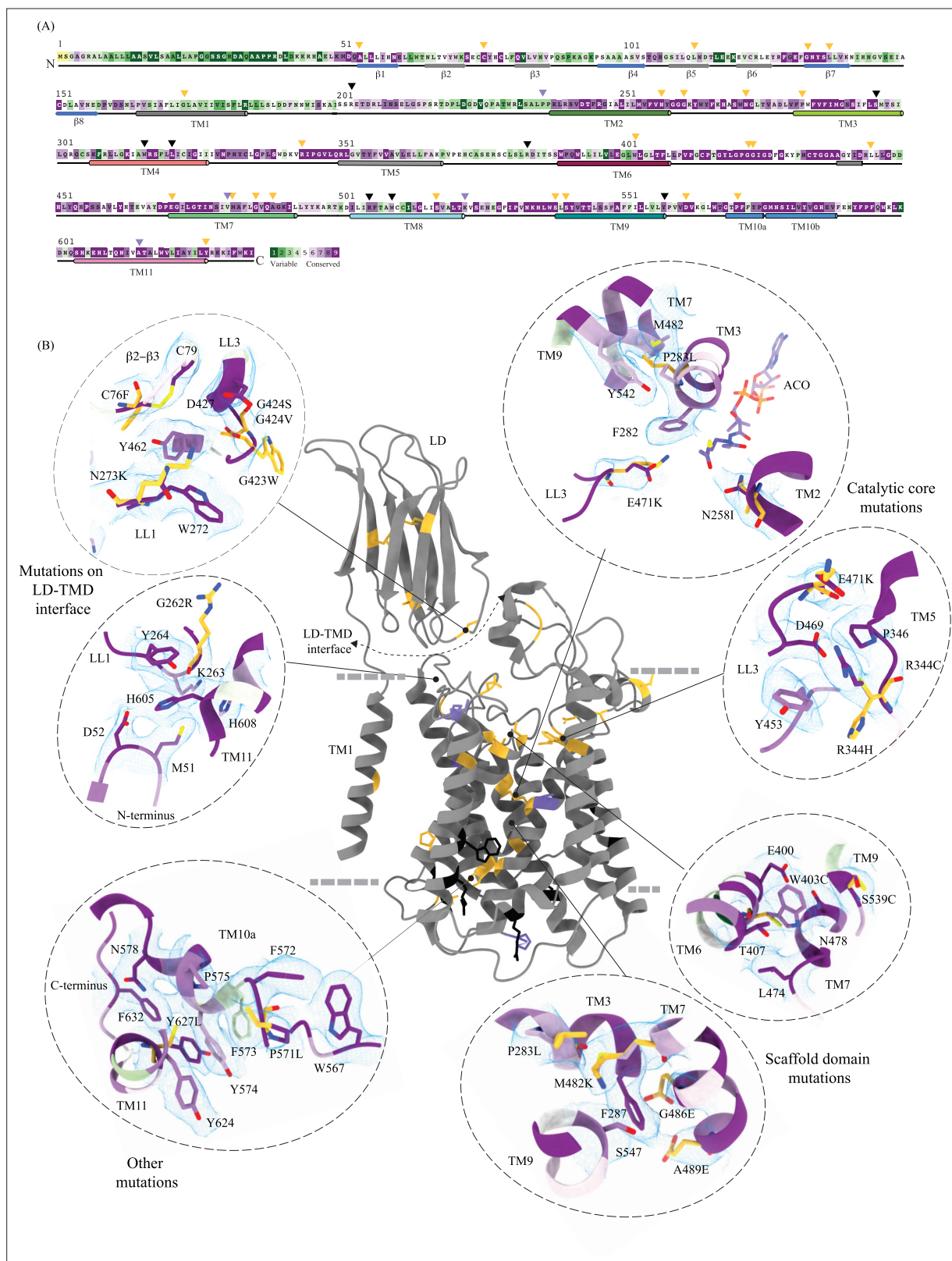


Figure 4. Molecular basis for mucopolysaccharidosis IIIC (MPS) IIIC mutation-induced dysfunction. **(A)** Evolutionary sequence conservation of heparan- α -glucosaminide N-acetyltransferase (HGSNAT). Amino acids are color-coded according to the conservation scores generated by the ConSurf web server using a Clustal multiple sequence alignment of homologs identified by PSI-BLAST (Ashkenazy et al., 2016). The positions of the mutations - missense (orange), nonsense (black), and polymorphisms (purple) - are indicated on the sequence by triangles. **(B)** MPS IIIC-causing mutations

Figure 4 continued on next page

Figure 4 continued

mapped on the HGSNAT structure. The color coding of the positions is the same as in panel A. Some of the missense mutants are highlighted in the insets (dashed ovals). We grouped them based on their position within the protein – LD-TMD interface, catalytic core, scaffold domain, and other C-terminal mutations. The insets show the 3D environment of the mutant sites on the wild-type HGSNAT color coded as per their evolutionary sequence conservation scores, and the potential disturbance to it caused by the mutation (orange side chains). The coordinates for mutant side chains were generated based on wild-type HGSNAT structure as input in FoldX webserver (**Schymkowitz et al., 2005**).

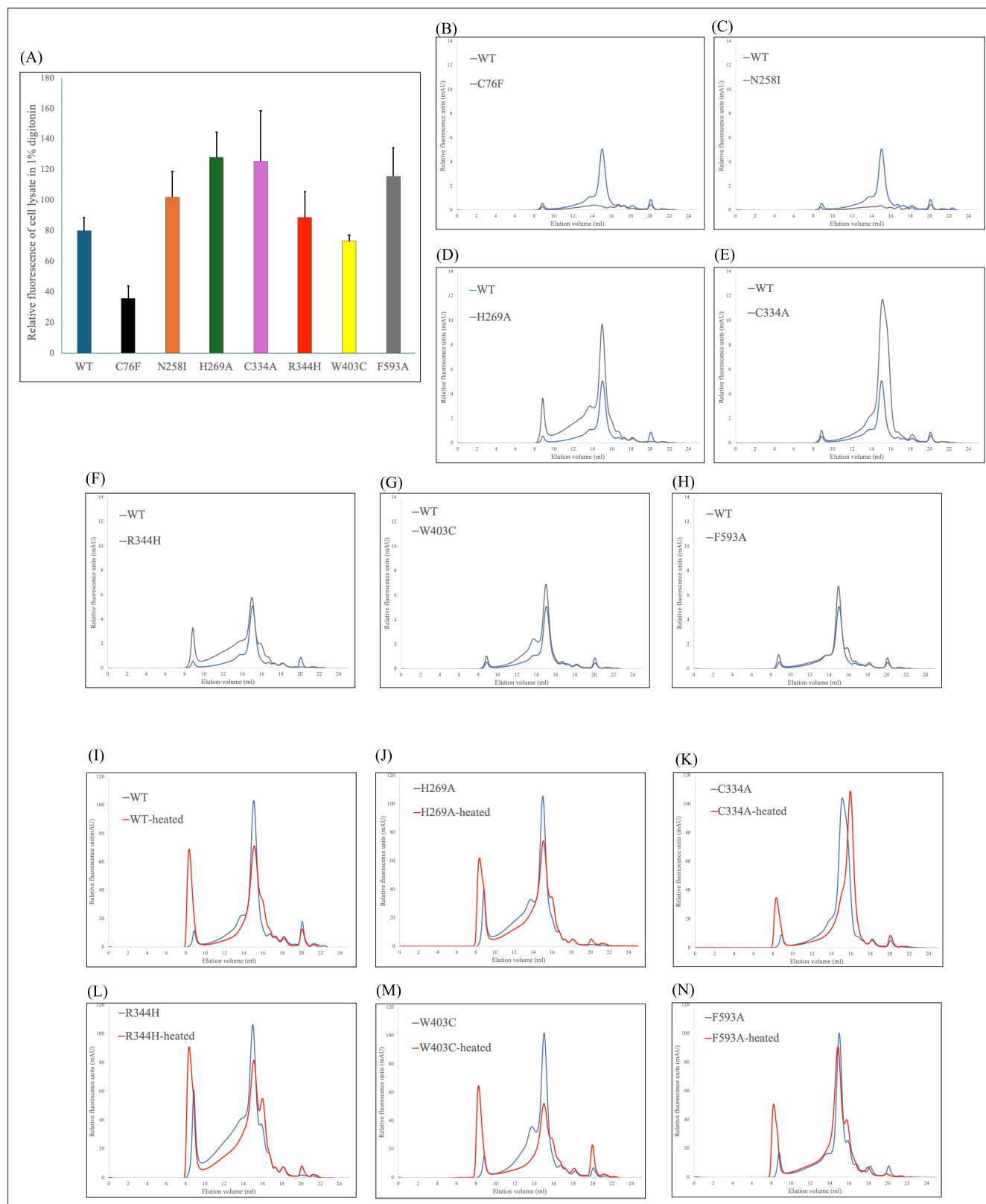


Figure 4—figure supplement 1. Expression and stability of heparan- α -glucosaminide N-acetyltransferase (HGSNAT) mutants. **(A)** A comparison of relative protein expression indicated by total GFP fluorescence in 100,000 HEK293S GnTI- cells expressing HGSNAT and its mutants. A representative of the experiment ($n=2$) is shown with error bars denoting standard errors of mean calculated from triplicate measurements. **(B–H)** A comparison of FSEC chromatograms of the HGSNAT mutants (gray chromatograms) with WT HGSNAT (blue chromatogram). C76F and N258I mutants show no peak

Figure 4—figure supplement 1 continued on next page

Figure 4—figure supplement 1 continued

at HGSNAT dimer position, and the remaining mutants' peak position is same as dimeric wild-type (WT) HGSNAT. **(I–N)** Relative stability of HGSNAT mutants analyzed by fluorescence-detection size-exclusion chromatography (FSEC). To estimate relative stability of mutants, the solubilized mutant cell lysates were heated at 65 °C for 15 min (red chromatograms) and the loss of HGSNAT peak in the resultant chromatograms were compared with non-heated samples (blue chromatograms). C334A, the mutant which breaks the disulfide at the dimer interface, results in a monomeric HGSNAT peak upon heating, while all other mutants retain their dimeric status.

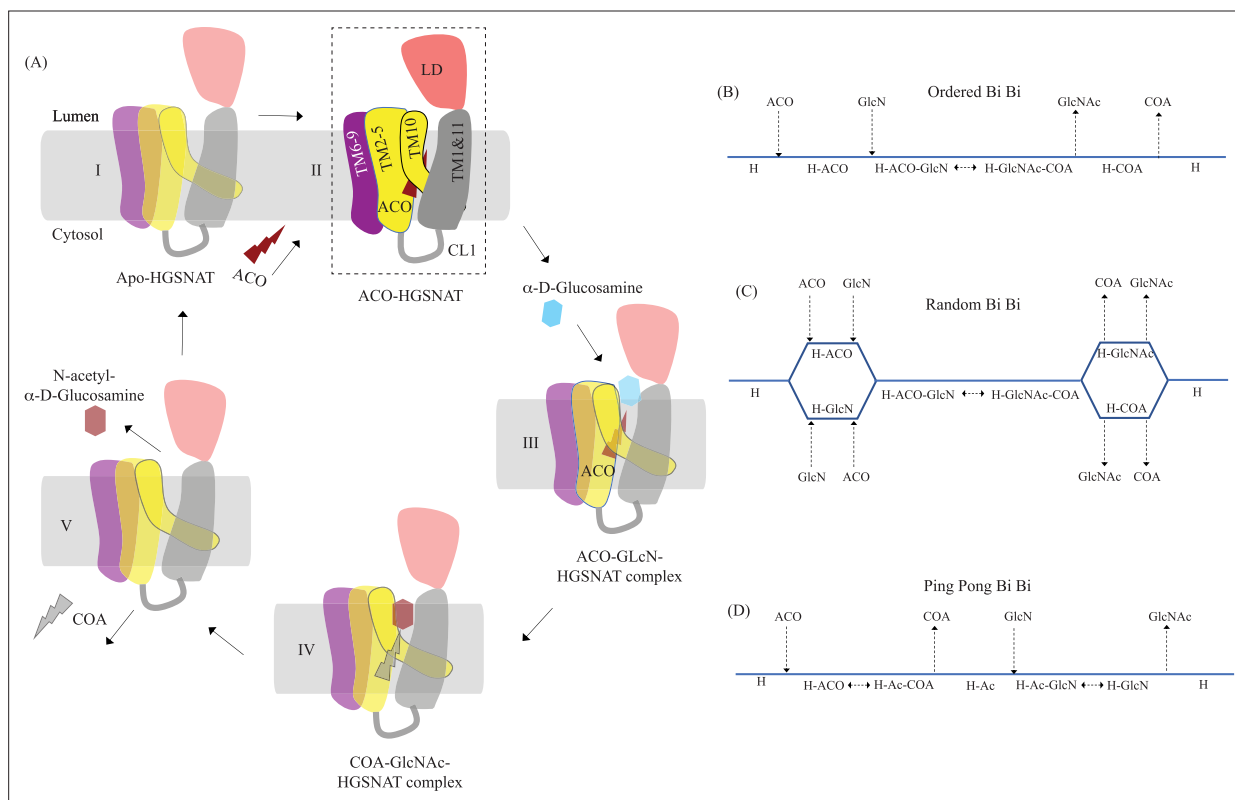


Figure 5. Proposed mechanism of acetyl transfer by heparan-α-glucosaminide N-acetyltransferase (HGSNAT). **(A)** HGSNAT (I) catalyzes a bisubstrate reaction of transferring acetyl group from cytosolic acetyl-CoA (ACO, red lightning) to terminal non-reducing α-D-Glucosamine (GlcN, blue hexagon) of luminal heparan sulfate (III and IV). After the acetyl group transfer, COA (gray lightning) and acetylated glucosamine (GlcNAc, red hexagon) are believed to be released to cytosol and lumen, respectively (V). Depending on the order of binding and release of substrates and products, enzyme-catalyzed bisubstrate reactions could either be sequential reactions (**B and C**) or ping pong reactions (**D**). The mechanism of reaction catalyzed by HGSNAT has been a longstanding debate. We believe that the acetyl-CoA bound HGSNAT structure presented in this work (II, dashed box) is in a cofactor primed conformation which could proceed by any of the bisubstrate reaction mechanisms shown in **B–D**. The function of luminal domain (LD) is unclear, and we believe it plays essential role in recognition of substrate and its positioning at the active site.

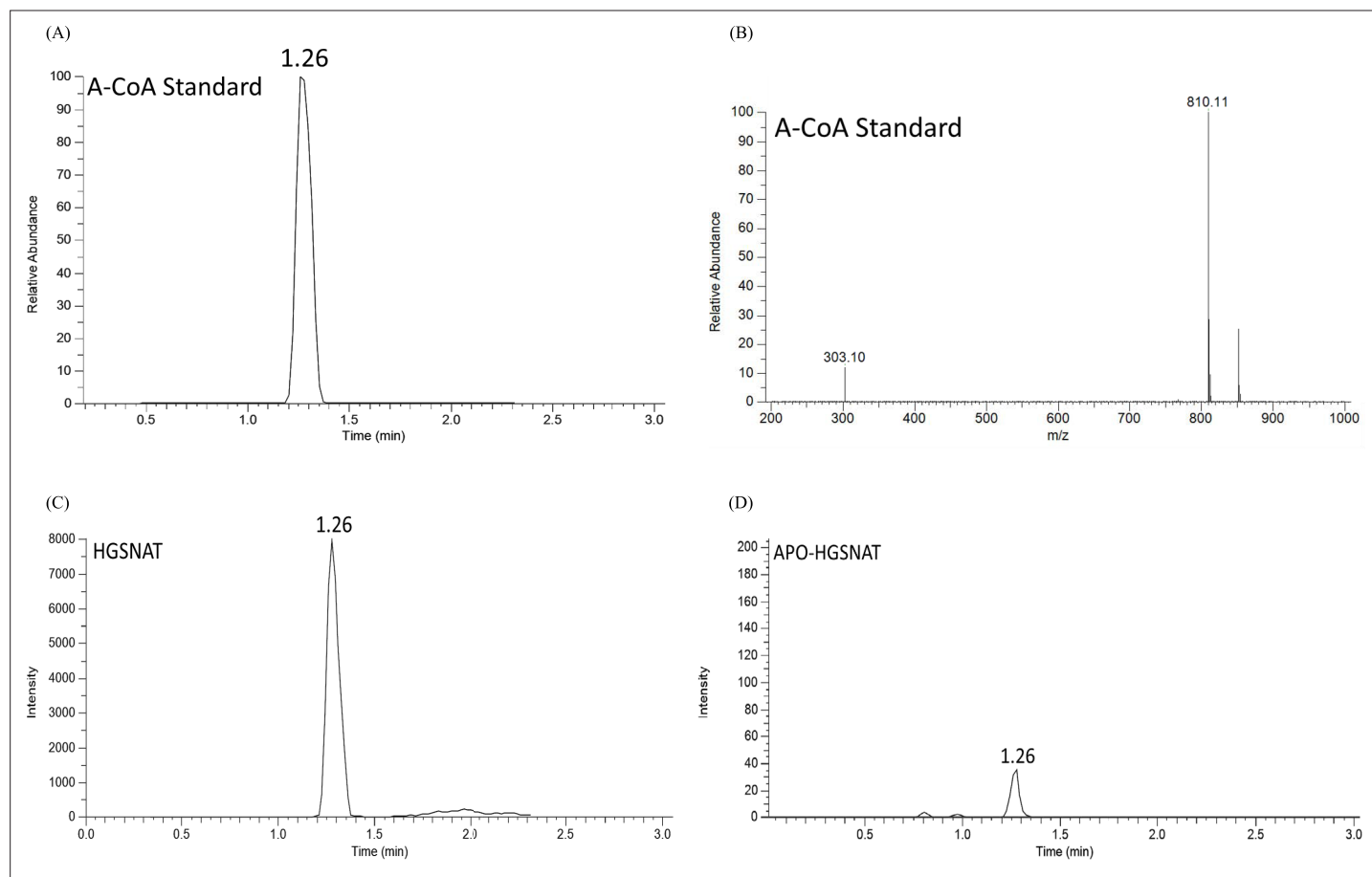


Figure 5—figure supplement 1. LC-MS analysis of purified heparan- α -glucosaminide N-acetyltransferase (HGSNAT). (A) LC profile and (B) MS/MS spectrum of acetyl-CoA (ACO) standard, showing the retention time (1.26 min), and precursor (810.1 m/z) and product (303.1 m/z) peaks in single reaction monitoring mode, respectively. (C) and (D) show relative LC peak intensities of endogenously bound ACO identified in purified HGSNAT before and after dialysis of the membranes, respectively.