Supplementary information for:

Molecular basis of collagen galactosylation by GLT25D1

Huanhuan Sun[1,2*], Min Zhang[2*#], Yunshu Shi[1], Di Zhang[1,2], Zhengkang Hua [1], Ping Yang[1,2], Yizheng Yang[1,2], Jiameng Li[1], Jiasheng Chen[3], Chao Wang[3], Peifen Li[2], Xinlin Hu[1], Yan Ke[1], Mingjie Zhang[4], Xiaotian Liu[4#], Hongjun Yu[1,5#]

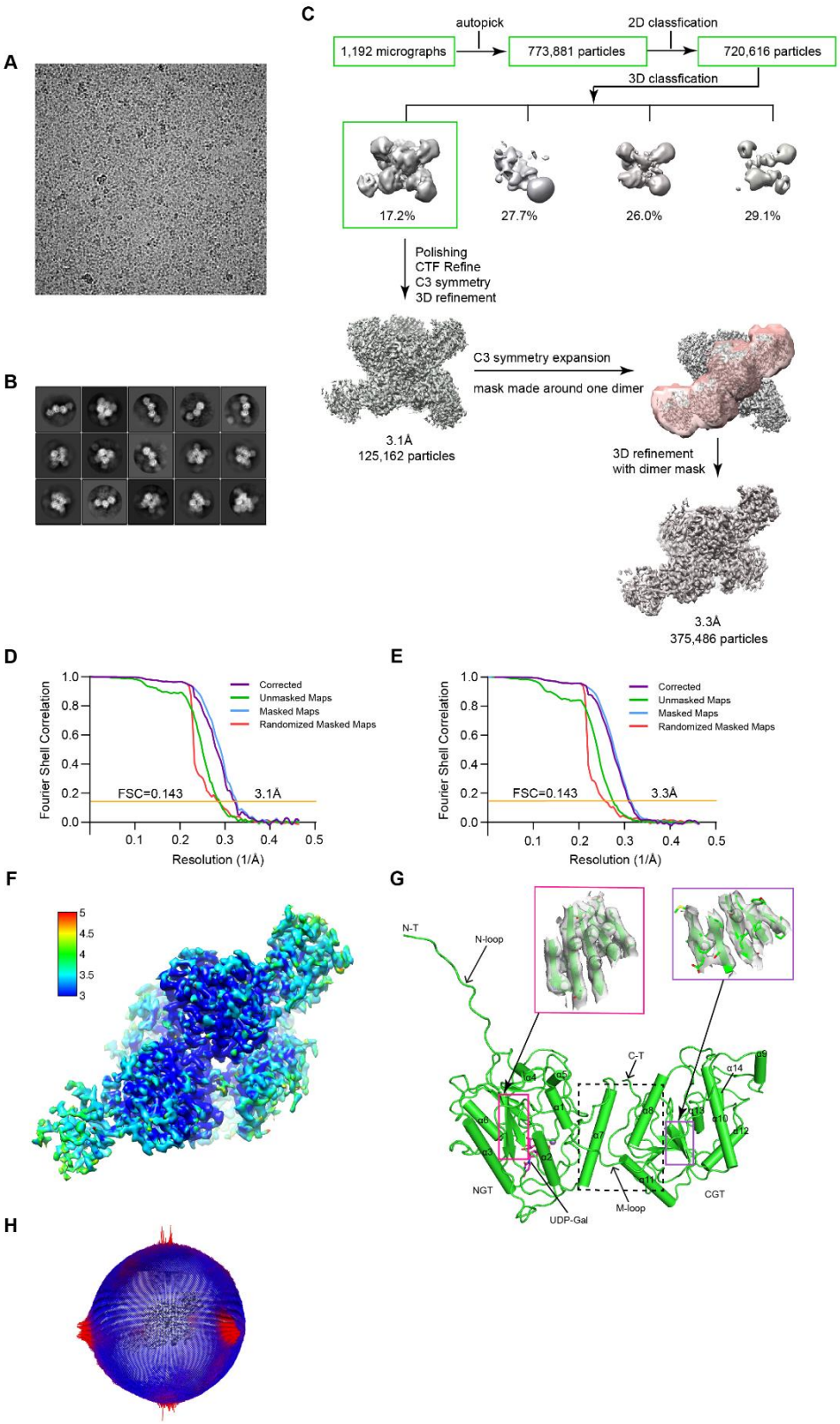
Correspondence to minzhang@hust.edu.cn or liuxt@sustech.edu.cn or hongjun_yu@hust.edu.cn

Supplementary items:

Supplementary Figs. 1-13

Supplementary Tables 1-4

SUPPLEMENTARY FIGURES

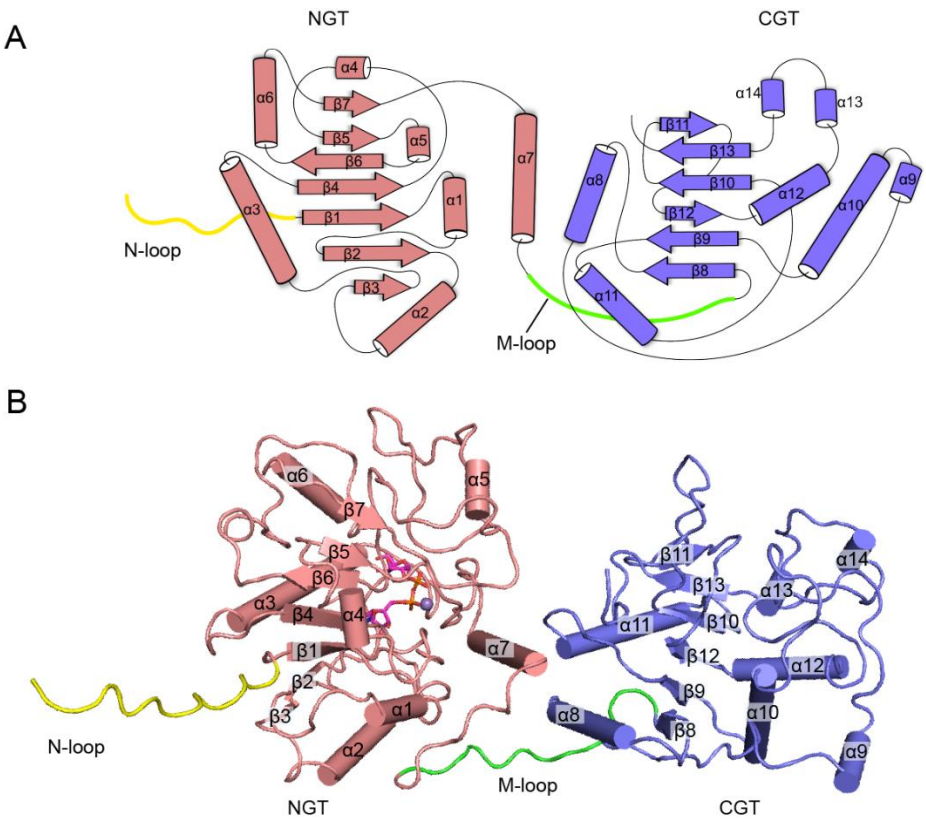


Supplementary Figure 1 | Cryo-EM analysis of GLT25D1.

A, Representative cryo-EM micrograph of GLT25D1. **B**, Representative 2D class averages.

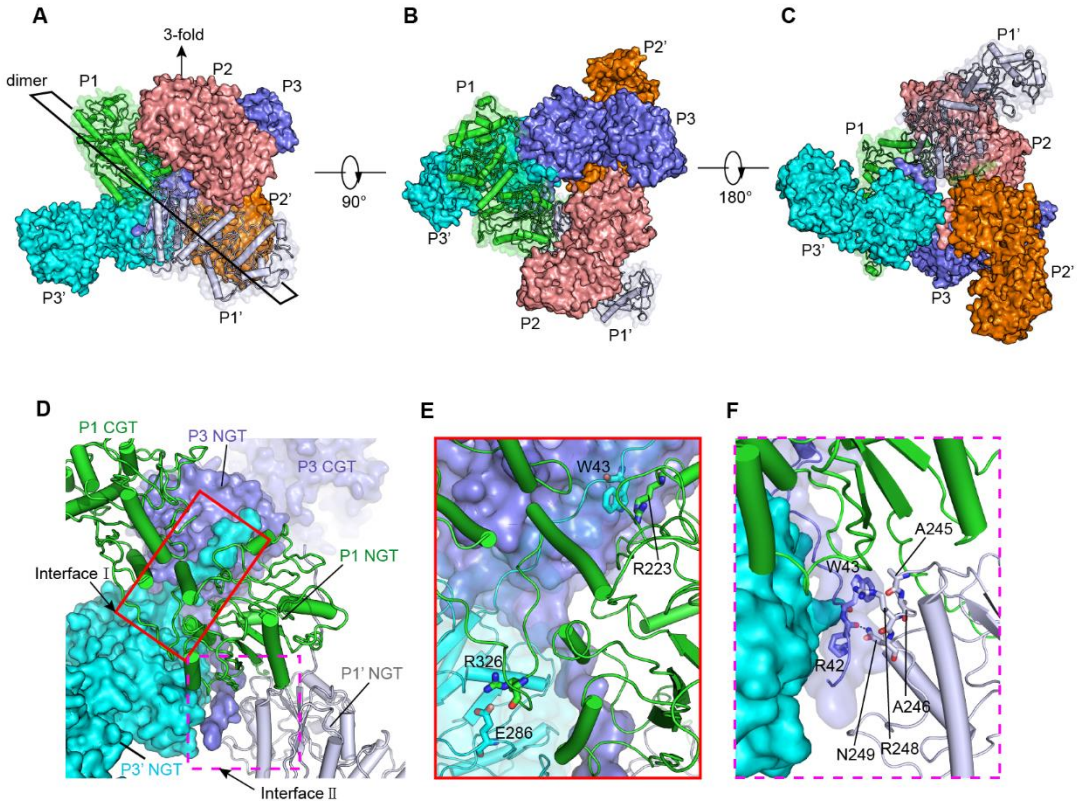
C, Flow chart outlining the cryo-EM data acquisition and processing pipeline for GLT25D1.

See Methods for more details. **D**, Gold-standard Fourier shell correlation (FSC) curves of the reconstructed map with C3 symmetry. **E**, Gold-standard Fourier shell correlation (FSC) curves of the reconstructed map after symmetry expansion and refinement using a dimer mask. **F**, Local resolution distribution of the cryo-EM map (contour level 0.022) of GLT25D1 after symmetry expansion and dimer mask refinement (corresponding to panel **E**), calculated using ResMap. **G**, Fit of cryo-EM map with GLT25D1 model in example regions: central β-sheets of the NGT (left) and CGT (right) domains. **H**, Angular distribution of all particles used in the final 3D reconstruction. For (**D**) and (**E**), source data are provided as a Source Data file.



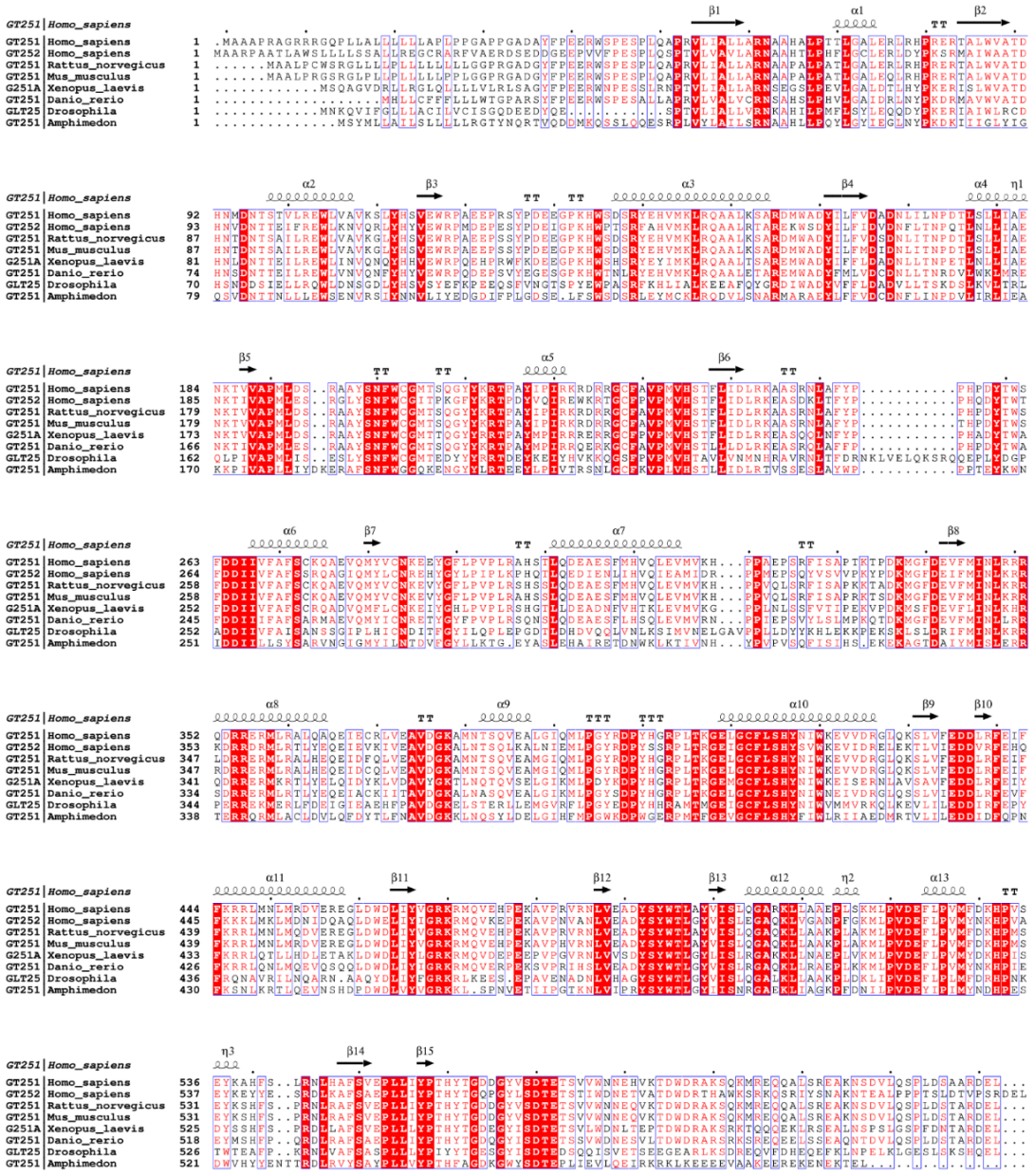
Supplementary Figure 2 | Bi-domain organization of GLT25D1.

A, Topological diagram of GLT25D1, with labeled secondary structural elements. The two GT domains, NGT and CGT, are colored in pink and blue, respectively. **B**, Overall structure of GLT25D1, with secondary structural elements labeled according to the color scheme in (**A**).



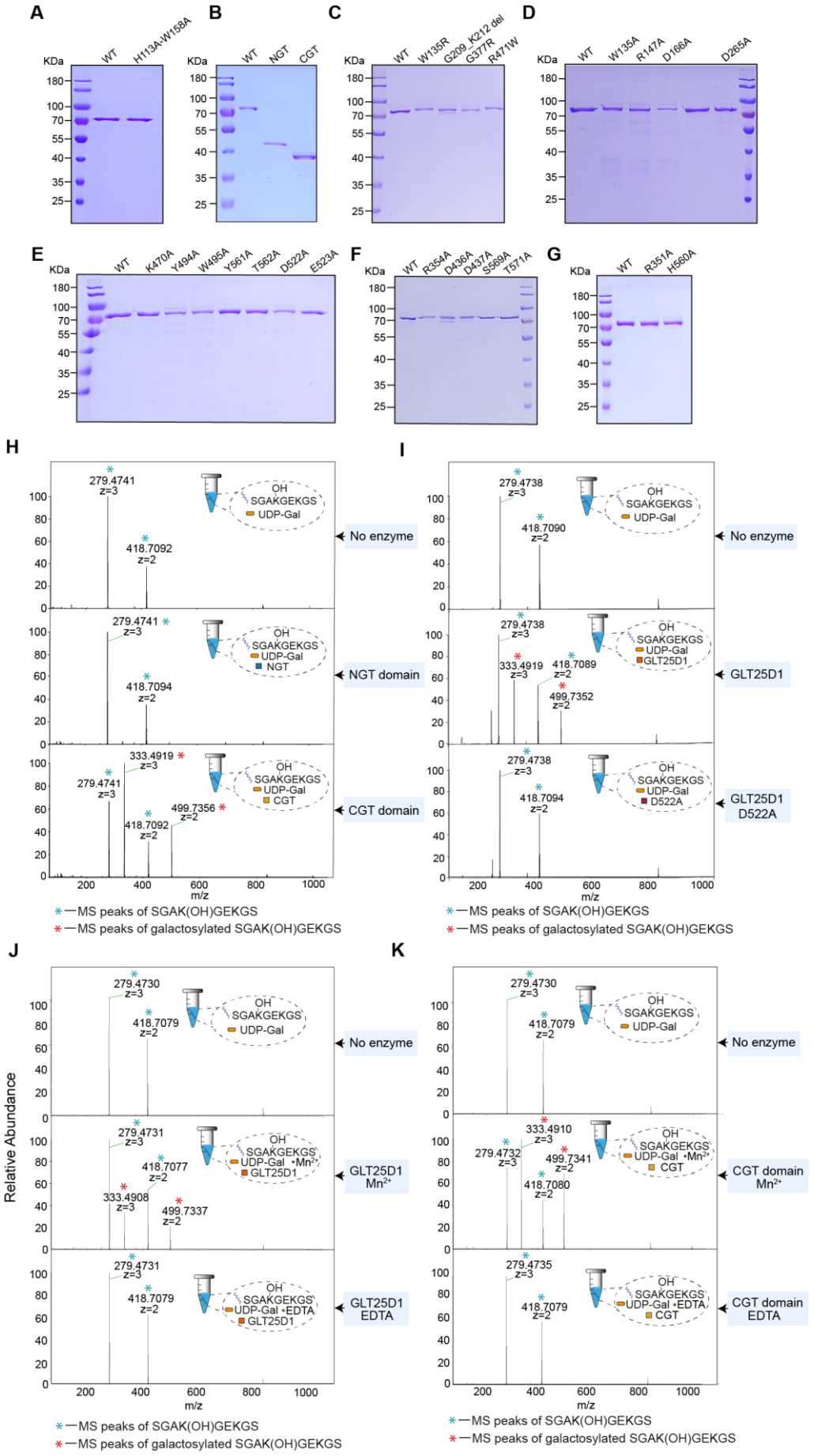
Supplementary Figure 3 | Assembly of GLT25D1 hexamer.

A-C, Structural views of hexameric GLT25D1 (**A**, side view; **B**, top view; **C**, bottom view). The hexamer is organized as a trimer of dimers, exhibiting C3 symmetry. The six composing protomers (P1-P3; P1'-P3') are depicted in different colors as labeled. The C3 symmetry axis is indicated by an arrow. **D**, GLT25D1 hexamer assembly is mediated by interactions between adjacent dimer units (e.g., P1-P1' and P3-P3'), forming two inter-dimer interacting interfaces (I and II) between their NGT domains. **E-F**, Close-up views of the inter-dimer interfaces highlighted in (**D**). **E**, solid boxed region; **F**, dashed boxed region. Interacting residues are shown as sticks.



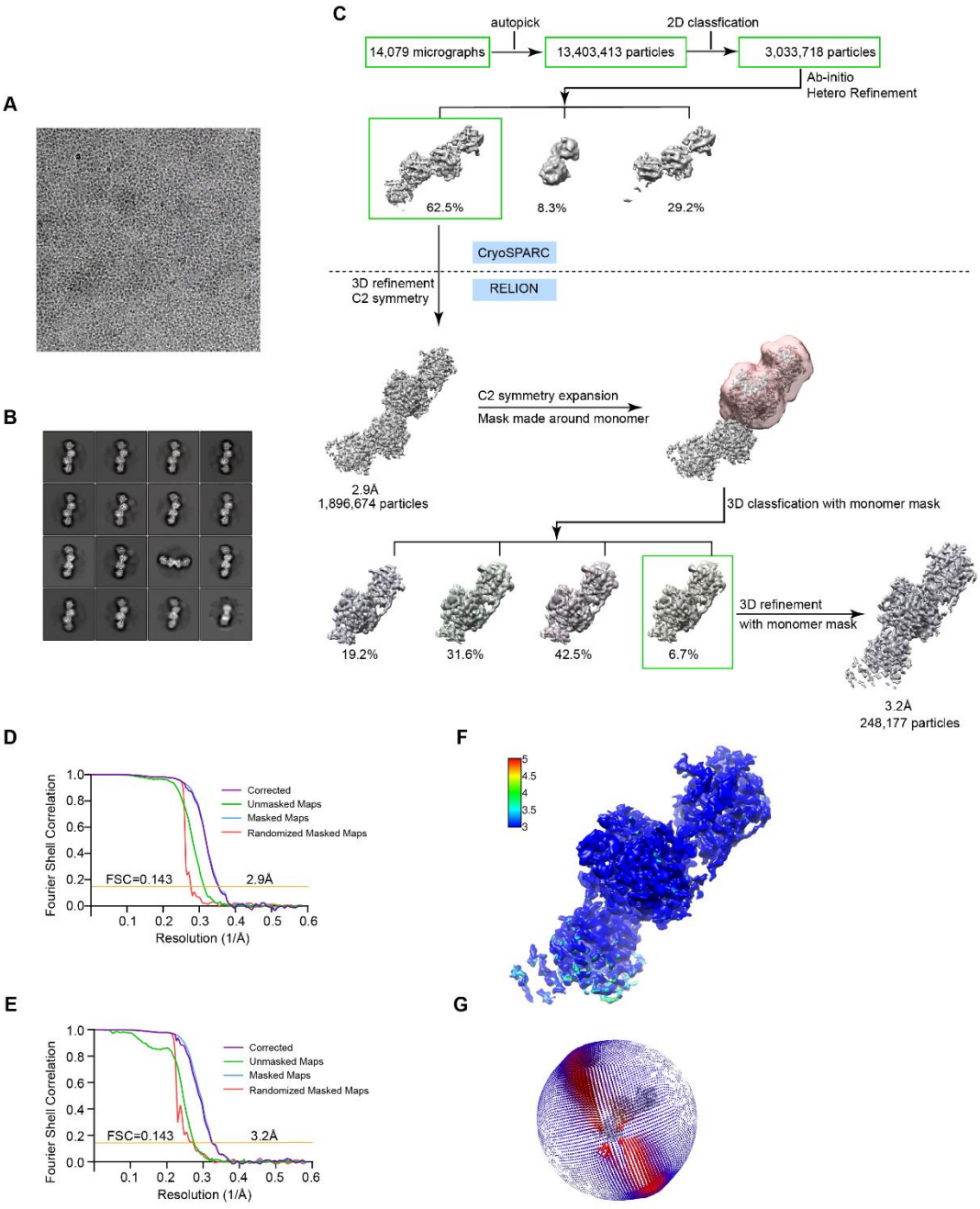
Supplementary Figure 4 | Multiple sequence alignment of GLT25D1 (procollagen galactosyltransferase) orthologs from representative metazoan species.

The primary sequence alignment was generated using Clustal Omega. Secondary structural element, key motifs and residues are annotated and labeled above the alignment. Protein sequences are from *Homo Sapiens* (GLT25D1, Q8NBJ5; GLT25D2, Q8IYK4), *Rattus norvegicus* (A6K9X0), *Mus musculus* (Q8K297), *Xenopus laevis* (A0JPH3), *Danio rerio* (A5PMF6), *Drosophila melanogaster* (Q8IPK4), and *Amphimedon queenslandica* (Sponge, A0AAN0JA54).


A
KDa
WT
H113A-W158A
180
100
70
55
40
35
25
B
KDa
WT
NGT
CGT
180
100
70
55
40
35
25
C
KDa
WT
W135R
G209_K212 del
G377R
R471W
180
100
70
55
40
35
25
D
KDa
WT
W135A
R147A
D166A
D265A
180
100
70
55
40
35
25
E
KDa
WT
K470A
Y494A
W495A
Y561A
T562A
D522A
E523A
180
100
70
55
40
35
25
F
KDa
WT
R354A
D436A
D437A
S569A
T571A
180
100
70
55
40
35
25
G
KDa
WT
R351A
H560A
180
100
70
55
40
35
25
H
279.4741
z=3
418.7092
z=2
OH
SGAKGEKGS
UDP-Gal
No enzyme
279.4741
z=3
418.7094
z=2
NGT
NGT domain
333.4919
z=3
279.4741
z=3
499.7356
z=2
418.7092
z=2
CGT
CGT domain
m/z
* — MS peaks of SGAK(OH)GEKGS
* — MS peaks of galactosylated SGAK(OH)GEKGS
I
279.4738
z=3
418.7090
z=2
No enzyme
279.4738
z=3
333.4919
z=3
418.7089
z=2
499.7352
z=2
GLT25D1
GLT25D1
279.4738
z=3
418.7094
z=2
D522A
GLT25D1
D522A
m/z
* — MS peaks of SGAK(OH)GEKGS
* — MS peaks of galactosylated SGAK(OH)GEKGS
J
Relative Abundance
279.4730
z=3
418.7079
z=2
No enzyme
279.4731
z=3
418.7077
z=2
333.4908
z=3
499.7337
z=2
GLT25D1
Mn2+
279.4731
z=3
418.7079
z=2
GLT25D1
EDTA
m/z
* — MS peaks of SGAK(OH)GEKGS
* — MS peaks of galactosylated SGAK(OH)GEKGS
K
279.4730
z=3
418.7079
z=2
No enzyme
333.4910
z=3
279.4732
z=3
499.7341
z=2
418.7080
z=2
CGT
CGT domain
Mn2+
279.4735
z=3
418.7079
z=2
CGT
CGT domain
EDTA
m/z
* — MS peaks of SGAK(OH)GEKGS
* — MS peaks of galactosylated SGAK(OH)GEKGS

Supplementary Figure 5 | Site-directed mutagenesis of GLT25D1 and mass spectrometric reaction product analysis of GLT25D1 mutants and truncations.

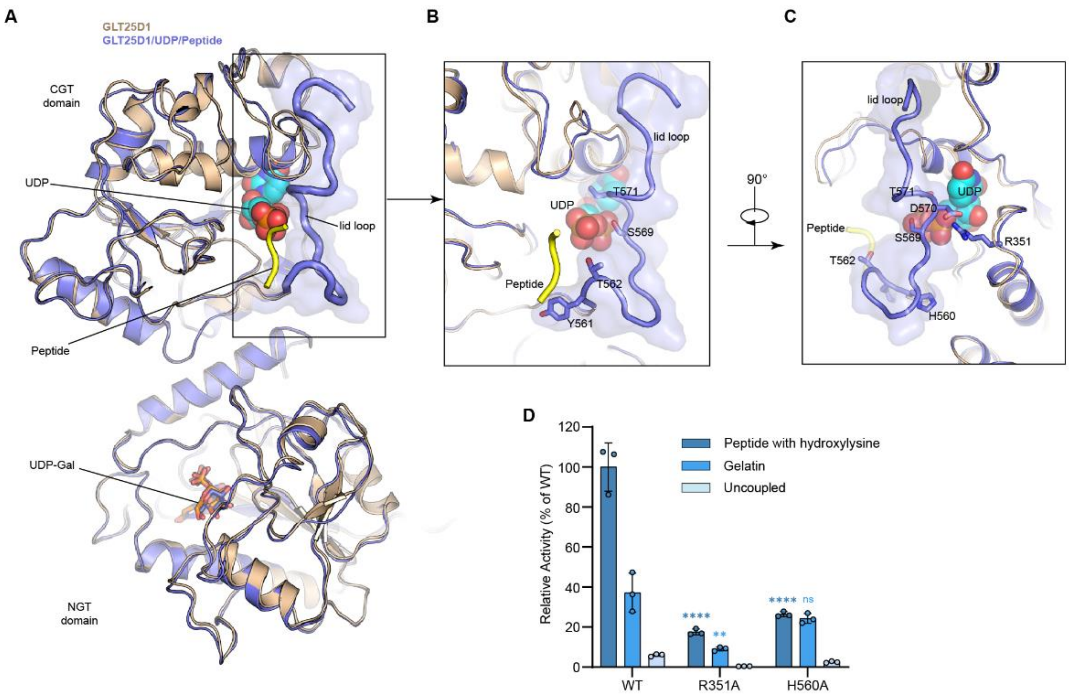
A-G, SDS-PAGE analyses of purified wild-type GLT25D1, GLT25D1 mutants, and truncations. **H-K**, *In vitro* activity of GLT25D1 variants under different conditions, assayed by mass spectrometry method. In each spectrum, unmodified peptides and galactosylated products are marked with blue (*) and red asterisks (*), respectively. **H**, MS reaction product analysis of GLT25D1 domain truncations (NGT and CGT domains). **I**, MS reaction product analysis of GLT25D1 and the D522A mutant. **J**, MS reaction product analysis of GLT25D1 in the presence of 100 μM Mn^{2+} (middle spectrum) or 1 mM EDTA (bottom spectrum). **K**, MS reaction product analysis of GLT25D1 CGT domain in the presence of 100 μM Mn^{2+} (middle spectrum) or 1 mM EDTA (bottom spectrum). Source data are provided as a Source Data file.



Supplementary Figure 6 | Cryo-EM analysis of ternary complex of GLT25D1 with substrates.

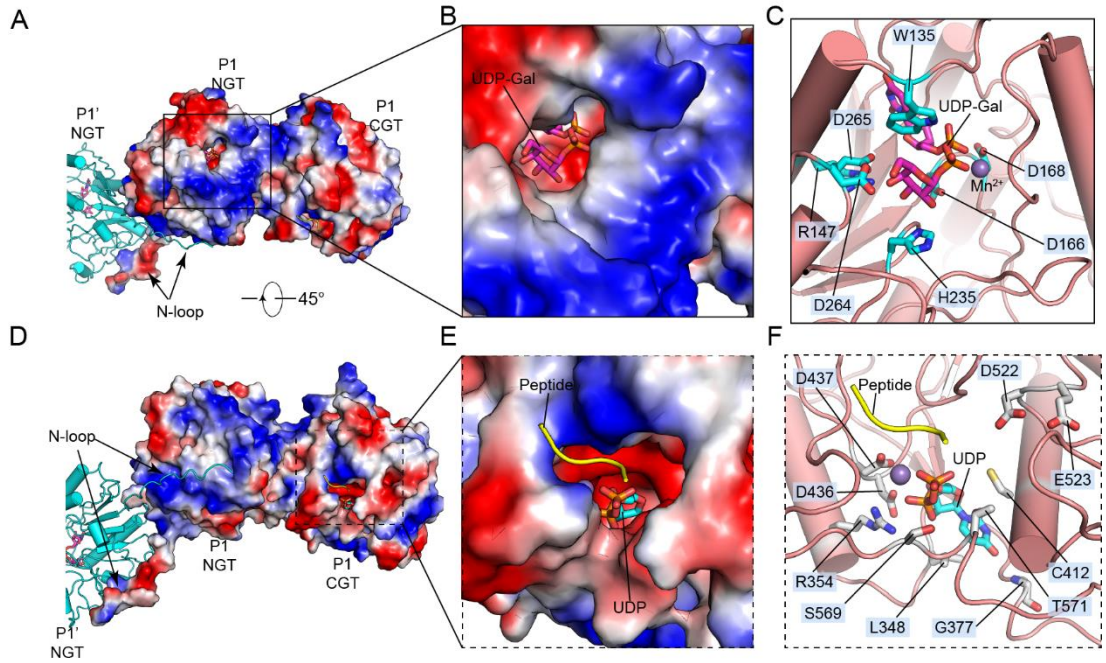
A, Representative cryo-EM micrograph of GLT25D1 ternary complex. **B**, Representative 2D class averages. **C**, Flow chart outlining the cryo-EM data acquisition and processing pipeline for GLT25D1 ternary complex. See Methods for more details. **D**, Gold-standard Fourier shell correlation (FSC) curves of the reconstructed map of GLT25D1 ternary complex dimer (refined with C2 symmetry). **E**, Gold-standard Fourier shell correlation (FSC) curves of the reconstructed map of GLT25D1 ternary complex after symmetry expansion and subsequent 3D classification and refinement using monomer mask. **F**, Local resolution

distribution of the cryo-EM map (contour level 0.022) of GLT25D1 ternary complex after symmetry expansion and subsequent 3D classification and refinement using monomer mask (corresponding to panel **E**), calculated using ResMap. **G**, Angular distribution of all particles used in the final 3D reconstruction. For (**D**) and (**E**), source data are provided as a Source Data file.



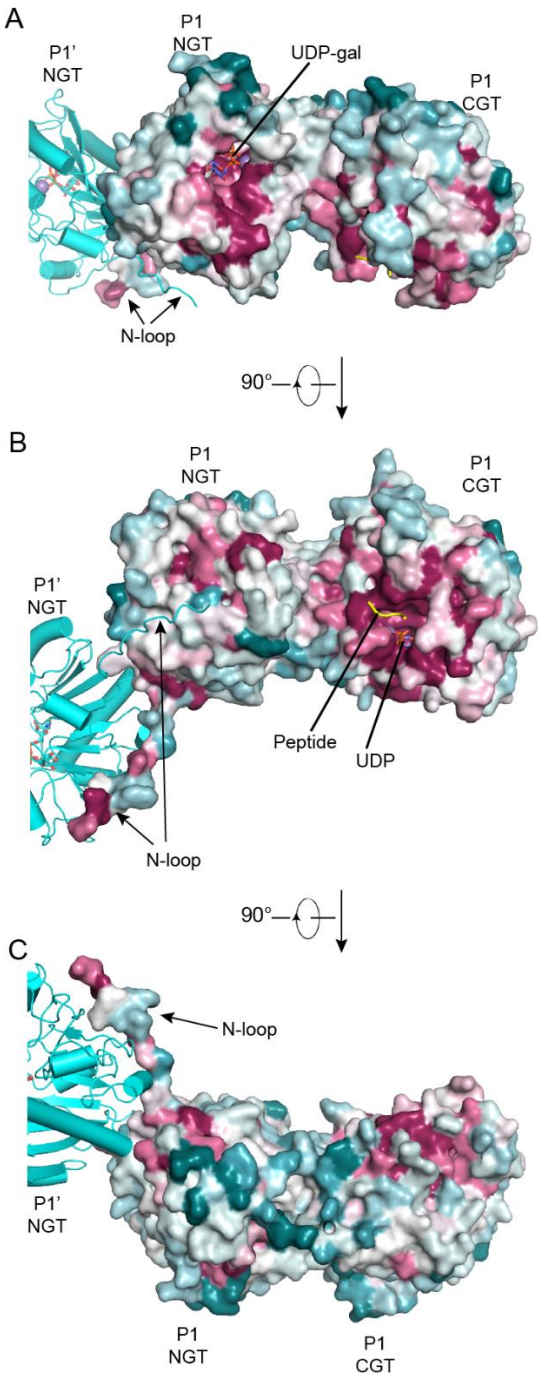
Supplementary Figure 7 | Conformational changes in GLT25D1 upon substrate binding.

A, Structural comparison between GLT25D1 (orange cartoon) and GLT25D1 complexed with UDP and peptide substrates (blue cartoon). A 20-residue loop (Y561-E580; shown as a thickened loop with transparent surface) adopts an ordered conformation upon substrate binding in the complex. The active site, where UDP (cyan stick) and peptide (yellow cartoon) bind, is marked by a black box. **B**, Close-up view of the boxed region in (**A**), highlighting the interaction between the lid loop and substrates. S569 and T571 interact with UDP, while Y561 and T562 interact with the peptide. **C**, Orthogonal view of the lid loop from (**B**), showing its stabilization by two conserved residues: H560 stack with the backbone of the lid loop, and R351 forms a salt bridge with D570. **D**, Relative activity of wild-type (WT) GLT25D1 and the mutants R351A and H560A, which disrupt lid loop stabilization as identified in (**C**). Activities were normalized to the amount of protein. Data are presented as mean ± s.d. for three independent replicates. Statistical significance was determined using ordinary one-way ANOVA, multiple comparison test (Dunnett's test at 95% CI) and differences are depicted as $^{**}p < 0.01$; $^{****}p < 0.0001$; ns, not significant. The exact p values are as follows. For the peptide with hydroxylysine substrate: WT vs R351A ($p < 0.0001$) and H560A ($p < 0.0001$). For the gelatin substrate: WT vs R351A ($p = 0.0019$) and H560A ($p = 0.0616$). Source data are provided as a Source Data file.



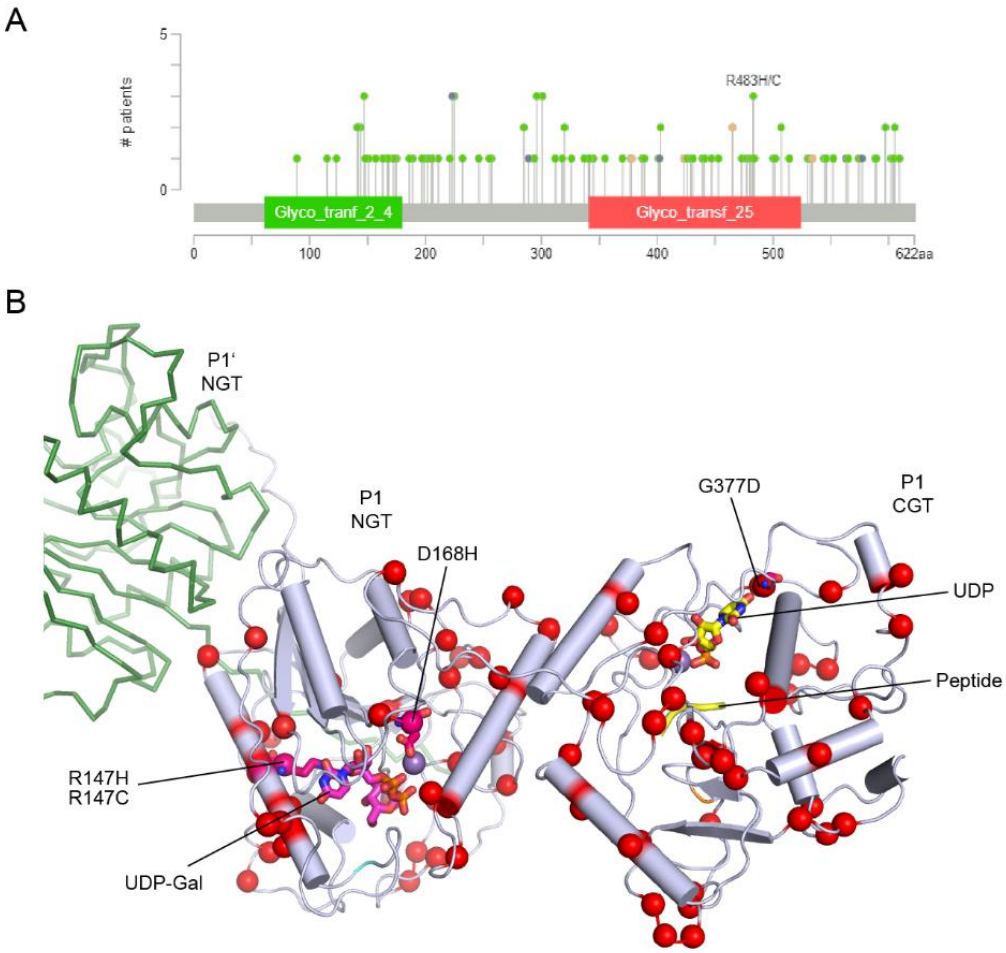
Supplementary Figure 8 | Two ligand binding sites on GLT25D1.

A, Electrostatic surface representation of a single protomer (P1) from GLT25D1 dimer. Blue: most positive; red: most negative. **B-C**, Close-up views of the ligand binding site within the NGT domain of GLT25D1. The ligand UDP-galactose is deeply embedded in the binding pocket (**B**), with specific residues involved in interactions with the galactose moiety shown as sticks (**C**). **D**, Electrostatic surface representation of protomer P1, rotated 45 degrees from the view in (**A**). **E-F**, Close-up views of the ligand binding site within the CGT domain of GLT25D1. The UDP and Mn^{2+} ion are coordinated by multiple conserved residues (shown as sticks) in the binding pocket (see **Supplementary Fig. 4** for residue conservation details).



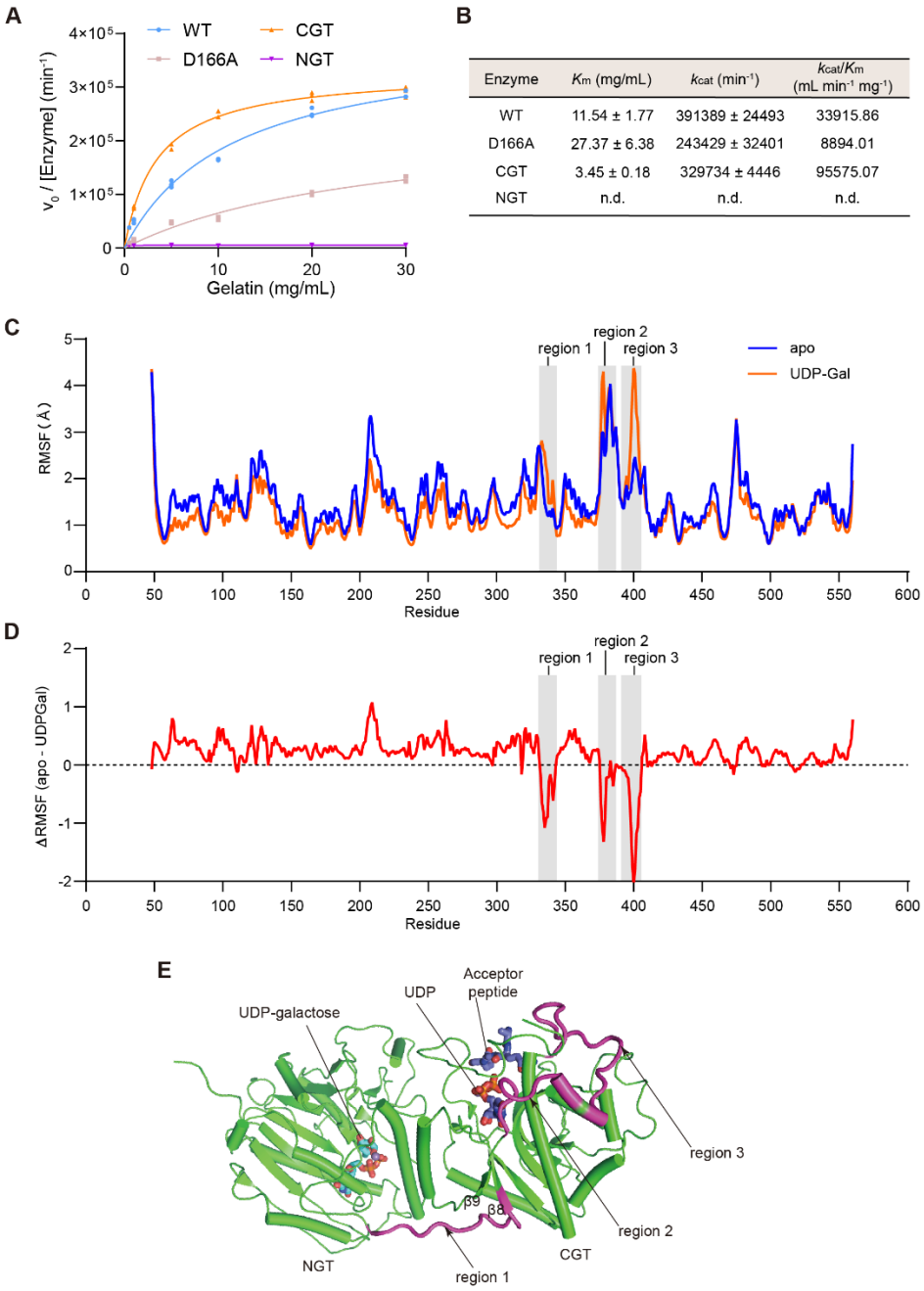
Supplementary Figure 9 | Surface conservation of metazoan GLT25D1.

A-C, Surface representation of human GLT25D1, colored based on sequence conservation (red: most conserved; green: least conserved). The surface conservation is highlighted on a single protomer (P1). The identified ligand binding sites on NGT and CGT domains are the most conserved regions on GLT25D1. GLT25D1 is rotated by 90 degrees in each panel (**A** to **C**) to illustrate the distribution of surface conservation.



Supplementary Figure 10 | Mapping of cancer-associated mutations onto the GLT25D1 structure.

A, Cancer-associated mutations in GLT25D1 were identified using cBioPortal (http://www.cbioportal.org/) and are represented as colored circles on the GLT25D1 sequence diagram. **B**, The cancer-associated mutations are mapped onto a single protomer (P1; shown as cartoons) of GLT25D1, with Cα atoms of the mutations depicted as red spheres. Four cancer-associated mutations (R147H, R147C, D168H, and G377D), situated at key ligand-binding interfaces, are highlighted as sticks.

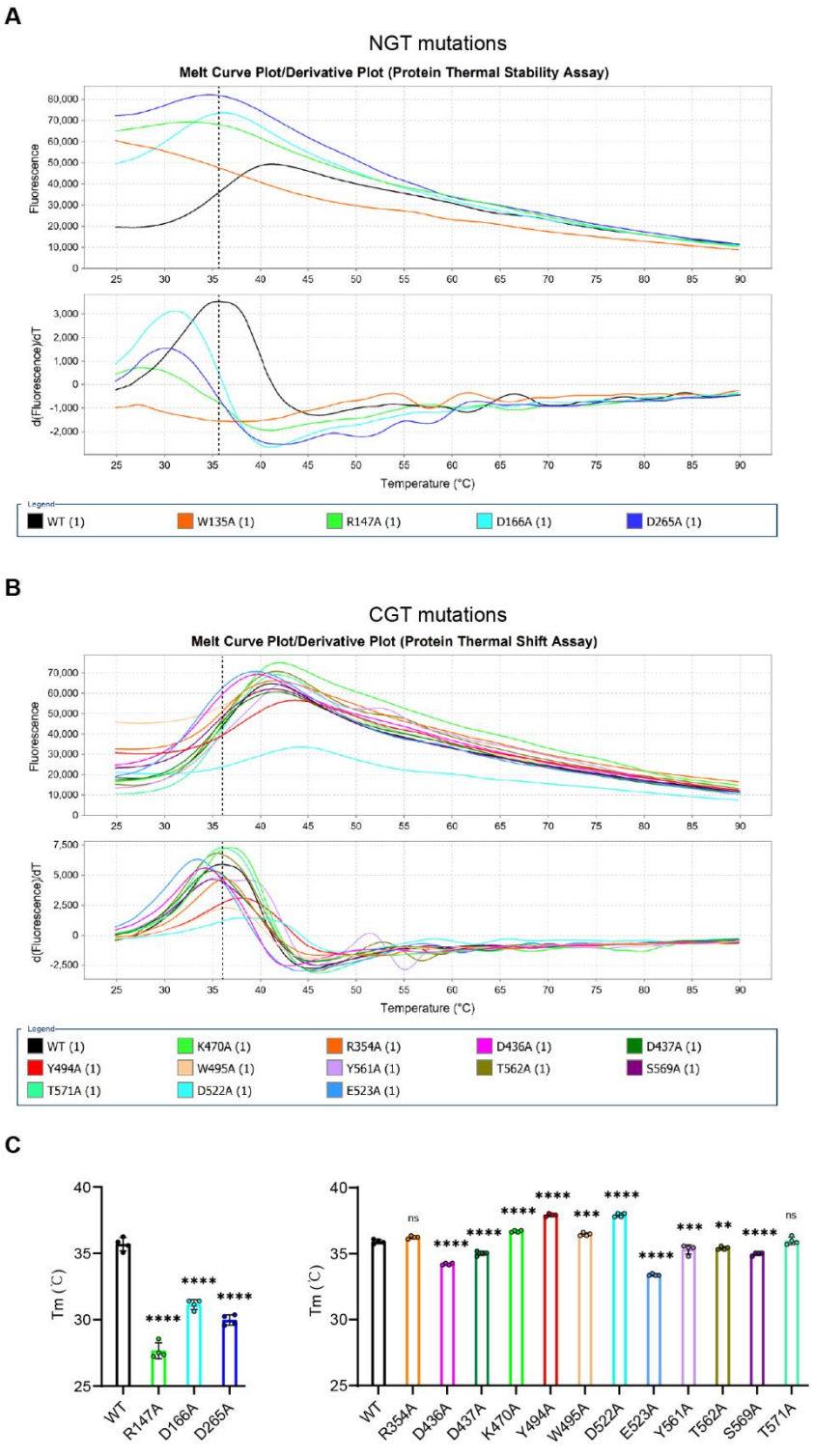


Enzyme	K_m (mg/mL)	k_{cat} (min^{-1})	k_{cat}/K_m ($mL\ min^{-1}\ mg^{-1}$)
WT	11.54 ± 1.77	391389 ± 24493	33915.86
D166A	27.37 ± 6.38	243429 ± 32401	8894.01
CGT	3.45 ± 0.18	329734 ± 4446	95575.07
NGT	n.d.	n.d.	n.d.

Supplementary Figure 11. Enzyme kinetics and molecular dynamics analysis of GLT25D1.

A, Enzyme kinetics of wild-type (WT) GLT25D1 and its variants, including the NGT mutant D166A, the isolated CGT domain, and the isolated NGT domain. Data are presented as mean ± s.d. from three independent replicates. Please refer to Methods section for experimental details. **B**, Enzyme kinetical parameters, corresponding to the analysis in (**A**). **C**, RMSF (root-mean-square fluctuation) curves from molecular dynamics simulations of apo-GLT25D1 (blue curve) and liganded GLT25D1 (with UDP-Galactose bound to NGT domain) (orange curve). **D**, Difference in RMSF (ΔRMSF) curve (red) between apo-GLT25D1 and liganded GLT25D1 (with UDP-galactose bound to NGT domain). Three

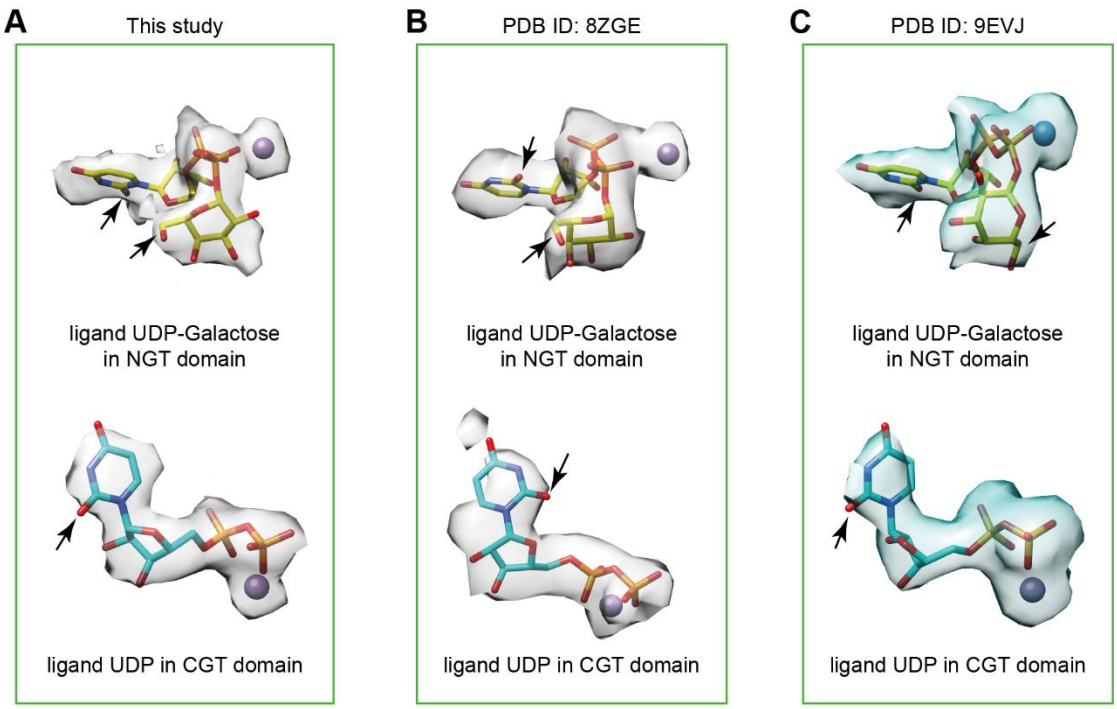
localized regions (regions 1-3) exhibiting increased RMSF upon ligand binding to the NGT domain are marked with shaded boxes in (**C**) and (**D**). **E**, Three regions (regions 1-3) identified in (**C**) and (**D**) are mapped to the ternary complex of GLT25D1, which contains UDP-Galactose in NGT domain, UDP and a hydroxylysine-containing acceptor peptide in CGT domain. Specifically, region 1 (residues 330–342) connects the NGT and CGT domains; regions 2 and 3 (residues 375–385 and 387–405, respectively) line the substrate-binding site in CGT domain. These three regions are spatially connected: β8 in region 1 runs parallel to β9 immediately preceding region 2, and regions 2 and 3 are directly linked. Source data are provided as a Source Data file.



Supplementary Figure 12 | Effect of GLT25D1 mutations on protein thermal stability.

A, Protein thermal shift assay of GLT25D1 wild-type (WT) and NGT domain mutants. Upper panel, the thermal denaturation profiles of analyzed proteins. Lower panel, the derivative plots of the thermal denaturation curves. **B**, Protein thermal shift assay of GLT25D1 wild-type (WT) and CGT domain mutants. Upper panel, the thermal denaturation profiles of analyzed proteins. Lower panel, the derivative plots of the thermal denaturation curves. **C**,

The melting temperatures (Tm) of wild-type (WT) and mutant proteins determined by protein thermal shift assays. Each bar represents the mean ± s.d. of four independent replicates. Statistical significance was determined using ordinary one-way ANOVA, multiple comparison test (Dunnett's test at 95% CI) and differences are depicted as $^{**}p < 0.01$; $^{***}p < 0.001$; $^{****}p < 0.0001$; ns, not significant. The exact p values are as follows. Left panel: WT vs R147A ($p < 0.0001$), D166A ($p < 0.0001$) and D265A ($p < 0.0001$). Right panel: WT vs R354A ($p = 0.0756$), D436A ($p < 0.0001$), D437A ($p < 0.0001$), K470A ($p < 0.0001$), Y494A ($p < 0.0001$), W495A ($p = 0.0004$), D522A ($p < 0.0001$), E523A ($p < 0.0001$), Y561A ($p = 0.0002$), T562A ($p = 0.0017$), S569A ($p < 0.0001$) and T571A ($p = 0.9993$). Notably, the Tm value for the W135A mutant could not be precisely determined due to its atypical profile; however, the observed curve clearly indicates markedly compromised protein stability. Source data are provided as a Source Data file.



Supplementary Figure 13. Comparison of nucleotide-ligand modeling and corresponding cryo-EM densities among three recent independent structural studies on GLT25D1.

A, GLT25D1 ternary complex with UDP/UDP-Galactose and a hydroxylysine-containing acceptor peptide (this study). The map shown as semi-transparent surface is contoured at level 0.022. **B**, GLT25D1/LH3 bound to UDP/UDP-galactose (Peng et al. *Nat Commun,* 16, 2436 (2025)). The map shown as semi-transparent surface is contoured at level 0.51. **C**, GLT25D1 bound to UDP/UDP-Galactose (De Marco et al. *Nat Commun,* 16: 3624-3624 (2025)). The map shown as semi-transparent surface is contoured at level 0.13. For each panel, the top section depicts the UDP-galactose bound to the NGT domain, and the bottom section shows the UDP bound to the CGT domain. Arrows highlight key modeling discrepancies among the corresponding ligands.

Supplementary Table 1 | Cryo-EM data collection, refinement and validation statistics

	Dimer (EMDB-63937) (PDB 9U7I)	Hexamer (EMDB-63938) (PDB 9U7J)	Ternary complex (EMDB-63936) (PDB 9U7H)
Data processing			
Magnification	130,000	130,000	130,000
Voltage (kV)	300	300	300
Electron exposure (e–/Å^2)	50	50	50
Defocus range (μm)	-1.2 to -3	-1.2 to -3	-1.2 to -3
Pixel size (Å)	1.076	1.076	0.855
Symmetry imposed	C1	C3	C1
Initial particle images (no.)	773,881	773,881	13,403,413
Final particle images (no.)	375,486	125,162	248,177
Map resolution (Å)	3.3	3.1	3.2
FSC threshold	0.143	0.143	0.143
Refinement			
Initial model used	AlphaFold3	This study	This study
Model resolution (Å)	3.30	3.31	3.28
FSC threshold	0.143	0.5	0.5
Map sharpening *B* factor (Å^2)	-124.7	-112.8	-117.1
Model composition			
Non-hydrogen atoms	8612	25836	4454
Protein residues	1042	3126	538
Ligands	2 UDP-galactose; 2 Mn^{2+}	6 UDP-galactose; 6 Mn^{2+}	1 UDP-galactose; 1 UDP; 2 Mn^{2+}
B factors (Å^2)			
Protein	78.2	78.2	56.3
Ligand	38.3	38.3	56.2
R.m.s. deviations			
Bond lengths (Å)	0.004	0.004	0.005
Bond angles (°)	0.820	0.820	0.856
Validation			
MolProbity score	2.07	2.09	2.16
Clashscore	10.75	11.33	11.68
Poor rotamers (%)	0.87	0.87	0.42
Ramachandran plot			
Favored (%)	91.07	91.07	88.72
Allowed (%)	8.73	8.73	11.09
Disallowed (%)	0.19	0.19	0.19

Supplementary Table 2 | The MS parameters of standards and ligand extract of GLT25D1

Analytes	RT (min)	ESI mode	Parent ion (m/z)	Daughter ion (m/z)	Cone voltage (V)	Collision energy(eV)
UDP-glucose	23.88	ESI-	564.8	78.89/322.9	100	46/22
UDP-galactose	25.33	ESI-	564.8	78.89/322.9	100	60/24
Protein extracts	25.33	ESI-	564.8	78.89/322.9	100	60/24

Supplementary Table 3 | A list of primers used for site-directed mutagenesis of GLT25D1

Mutation	Sequence (5'→3')	
H113A	Forward	TCCGTGGAGTGGCGGCCAGCAGA
	Reverse	TGCGTACAAACTCTTCACGGCCACCAGCCAC
W158A	Forward	GCTGATTACATCCTGTTTGTAGATGCGG
	Reverse	GGCCATGTCTCGAGCTGATTTCAGGGCTG
R61A	Forward	AACGCGGCCCACGCGTTGCC
	Reverse	TGCCGCCAACAGCGCGATGAGCACGC
V143A	Forward	GCCATGAAGTTGCGCCAGGCAGCC
	Reverse	GTCTGACTCACGCTACGAGCAT
D166A	Forward	GCCGCGGACAACCTGATCCTCAACC
	Reverse	TACAAACAGGATGTAATCAGCC
D168A	Forward	AACCTGATCCTCAACCCTGACACAC
	Reverse	AGCCGCATCTACAAACAGGATGTAATCAGC
W135A	Forward	TCTGACTCACGCTACGAGCATGTC
	Reverse	TGCGTGTTTCGGGCCTTCCTCGTCCG
R147A	Forward	CAGGCAGCCCTGAAATCAGCTCG
	Reverse	AGCCAACTTCATGACATGCTCGTAGCG
H235A	Forward	TCGACCTTCCTGATCGACCTGCG
	Reverse	TGCCACCATGGGAACTGCAAAGCAGC
D265A	Forward	ATCATCGTCTTTGCCTTCTCCTGCAAG
	Reverse	AGCGTCAAAGGACCAGGTGTAGTCAGG
R354A	Forward	GCTCGGGAGCGCATGCTGCG
	Reverse	GTCCTGCCGCCGCCTCAGG
D436A	Forward	GACCTGCGTTTTGAGATCTTCTTC
	Reverse	AGCCTCAAACACAAGCGATTTCTGC
D437A	Forward	GCCCTGCGTTTTGAGATCTTCTTC
	Reverse	ATCCTCAAACACAAGCGATTTCTGC
S569A	Forward	GACACCGAGACCTCAGTCGTATG
	Reverse	AGCCACATAGCCATCGTCTCCTG
T571A	Forward	GCCGAGACCTCAGTCGTATGGAACA
	Reverse	GTCACTCACATAGCCATCGTCTCCTG
K470A	Forward	CGGATGCAGGTGGAGCACCCCGAG
	Reverse	TGCCCGGCCCACATAGATGAGGTC
Y494A	Forward	GCCTGGACCCTGGCCTACGTGATC
	Reverse	GGAATAGTCGGCCTCCACCAGG
W495A	Forward	GCCACCCTGGCCTACGTGATCTCC
	Reverse	GTAGGAATAGTCGGCCTCCACCAGGTTC
Y561A	Forward	GCCACAGGAGACGATGGCTATGTG
	Reverse	GTGTGTGGGGTAGATGAGCAGC
T562A	Forward	GCAGGAGACGATGGCTATGTGAG
	Reverse	GTAGTGTGTGGGGTAGATGAGC
D522A	Forward	CCGAGTTCCTGCCCGTCATGTTCGAC

	Reverse	CCACAGGCAGCATCTTGGAGAGCG
E523A	Forward	GCATTCCTGCCCGTCATGTTCGAC
	Reverse	GTCCACAGGCAGCATCTTGGAG
R351A	Forward	CTCAGGACCGGCGGGAGCGCATGC
	Reverse	CCCGCCTCAGGTTGATCATGAAGACC
H560A	Forward	GCCTACACAGGAGACGATGGC
	Reverse	TGTGGGGTAGATGAGCAGCGGC
L151R	Forward	AAATCAGCTCGAGACATGTGGGCTG
	Reverse	CCTGGCTGCCTGGCGCAACTTCATGAC
A154P	Forward	CGAGACATGTGGGCTGATTACATCC
	Reverse	AGGTGATTTCAGGGCTGCCTGGCGCA
W135R	Forward	GTCTGACTCACGCTACGAGCATGTC
	Reverse	CGGTGTTTCGGGCCTTCCTCGTCCG
G209_K212del	Forward	CGCACACCTGCCTACATCCCTATC
	Reverse	CTGGGAAGTCATTCCACACCAGAAG
G377R	Forward	AAAGCCATGAACACCAGCCAGGTG
	Reverse	CCTGTCCACGGCCTCCACCAGCCGG
R471W	Forward	TGGATGCAGGTGGAGCACCCCGAG
	Reverse	CTTCCGGCCCACATAGATGAGGTC

Supplementary Table 4 | System composition for MD simulations

	GLT25D1 apo	GLT25D1 with ligand in NGT
Box dimensions	131.60 × 131.60 × 131.60 Å^3	131.49 × 131.49 × 131.49 Å^3
Total number of atoms	104859	104508
Number of water molecules	32053	31923
Salt concentration	150 mM NaCl	150 mM NaCl