Supporting Information for:

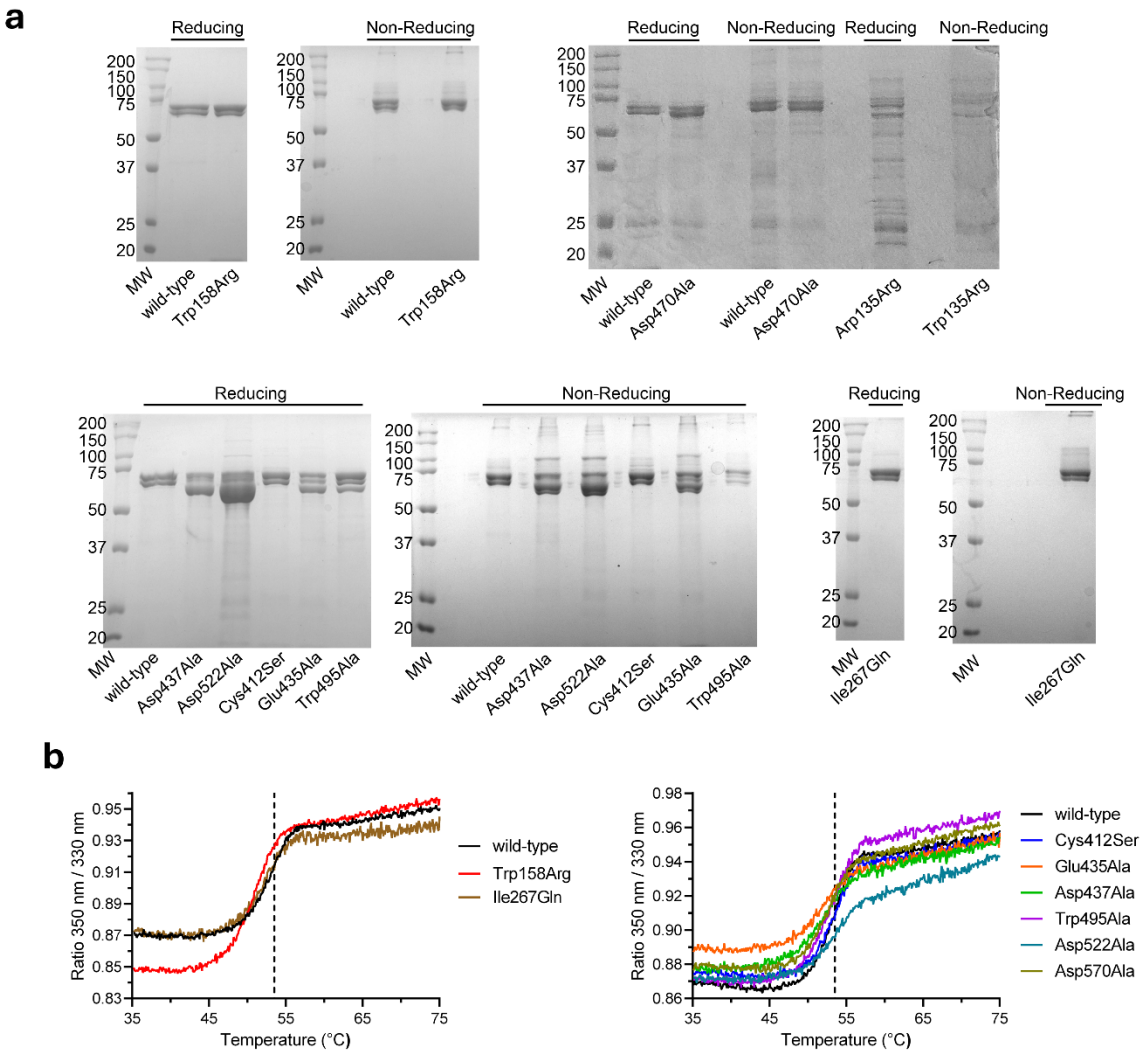
Molecular Structure and Enzymatic Mechanism of the Human Collagen Hydroxylysine Galactosyltransferase GLT25D1/COLGALT1

by Matteo De Marco, Sristi Raj Rai, Luigi Scietti, et al.

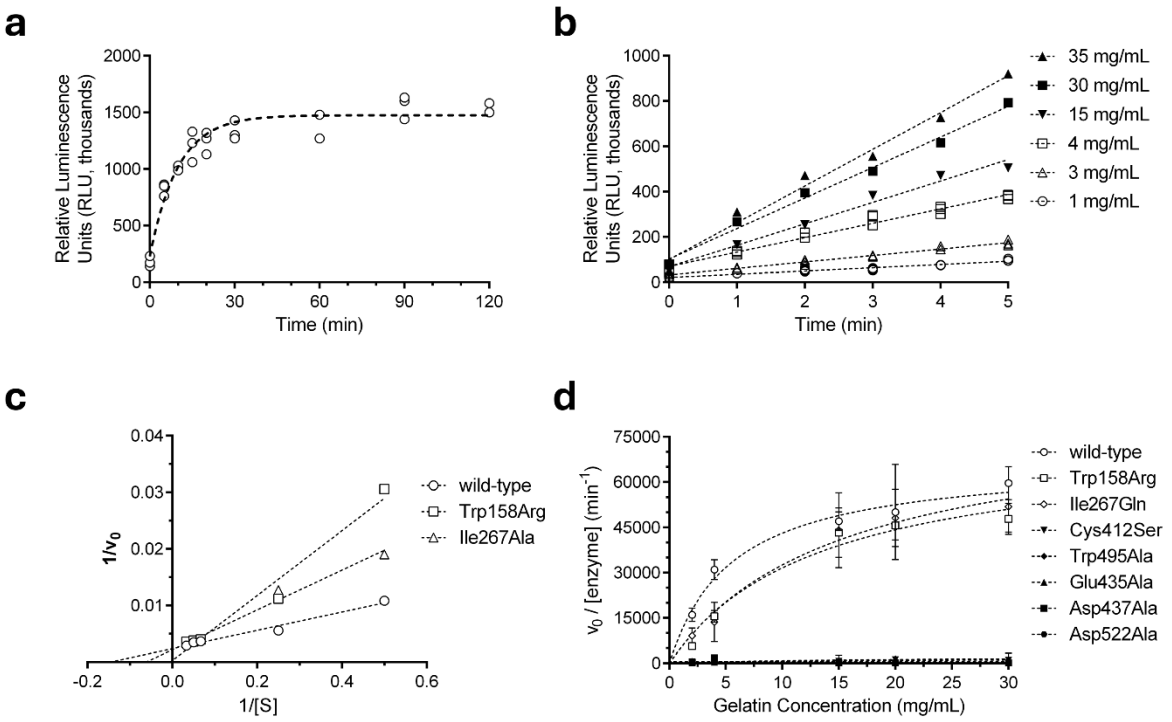
Correspondence to: Federico Forneris, e-mail: federico.forneris@unipv.it

Table of Contents:

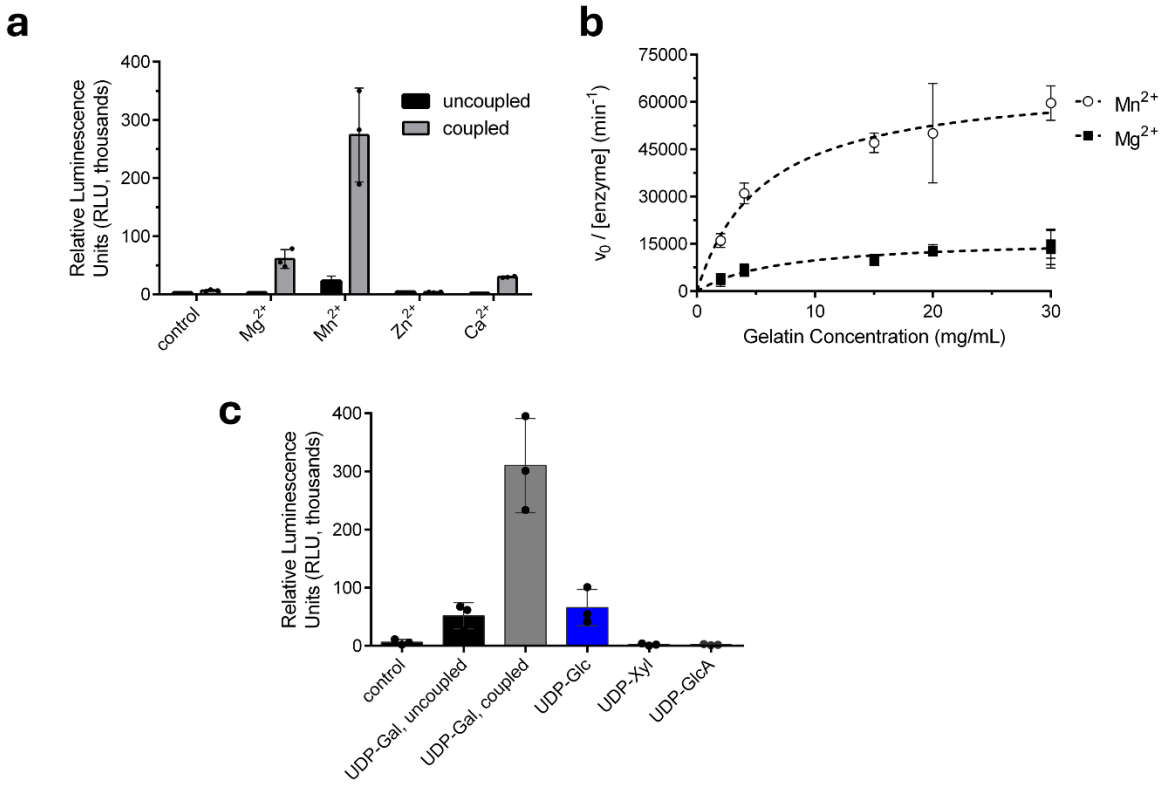
SUPPLEMENTARY FIGURES



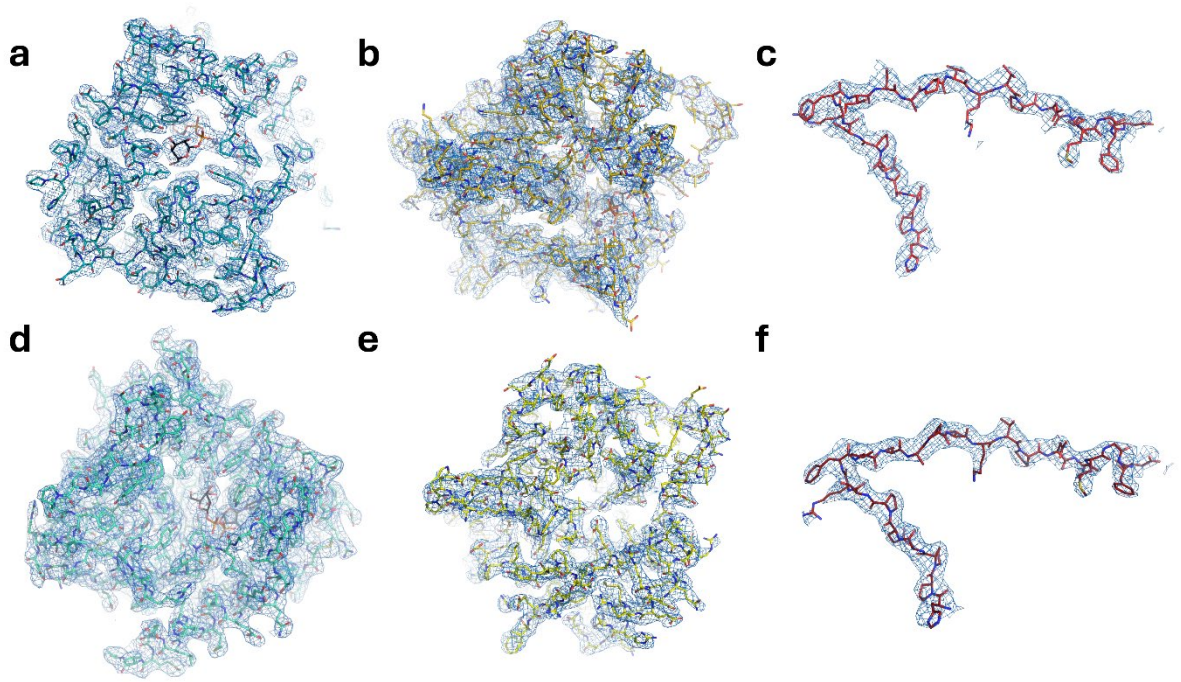
Supplementary Figure 1. Quality assessment of enzyme sample preparations. (a) All wild-type and mutant GLT25D1/COLGALT1 that could be generated underwent testing for correct folding using reducing and non-reducing SDS-PAGE. MW, molecular weight marker. (b) DSF analysis of the recombinant wild-type and mutant GLT25D1/COLGALT1 samples shown in panel (a). Mutant Trp135Arg was not tested due to the poor sample quality. The dashed vertical line indicates the unfolding temperature of wild-type GLT25D1/COLGALT1.



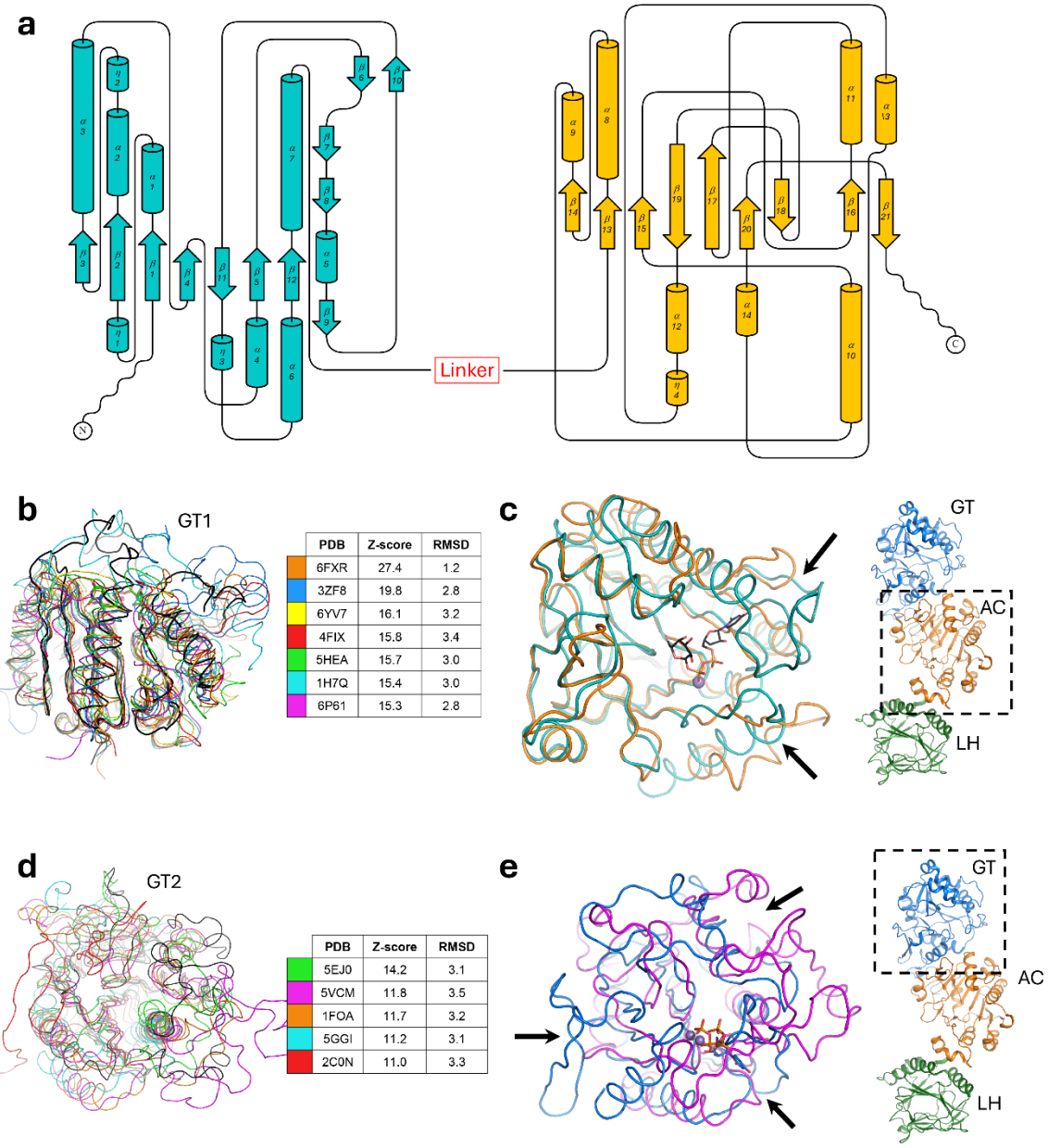
Supplementary Figure 2. Development of a luminescence-based steady-state kinetics assay to evaluate GLT25D1/COLGALT1 enzymatic activity. (a) Individual time points obtained from endpoint measurements using 1 mg/mL GLT25D1/COLGALT1 and 4 mg/mL gelatin substrate displayed exponential profiles, as confirmed by the nonlinear fit performed using *Graphpad Prism*[1], enabling identification of the initial linearity region calculation of initial velocities. (b) Linear fitting of time points collected within the first five minutes of the enzymatic reactions performed in presence of different concentrations of gelatin (represented by different symbols as reported in the figure legend) at constant enzyme concentration (i.e., 1 mg/mL) allows determination of initial velocities. (c) double-reciprocal plot showing the results obtained for three different GLT25D1/COLGALT1 variants (constant enzyme concentration 1 mg/mL) shows linearity, enabling quantitation of K_M and v_{max} values. (d) Results of the steady-state kinetics analysis of GLT25D1/COLGALT1 wild-type and mutants subject to this study. Error bars represent standard deviations from average of triplicate independent experiments. Initial velocity values, expressed as apparent turnover numbers (v_0 / [enzyme]) were subject to nonlinear fit using the Michaelis-Menten equation in *Graphpad Prism*[1]. Results are shown in Supplementary Table 5.



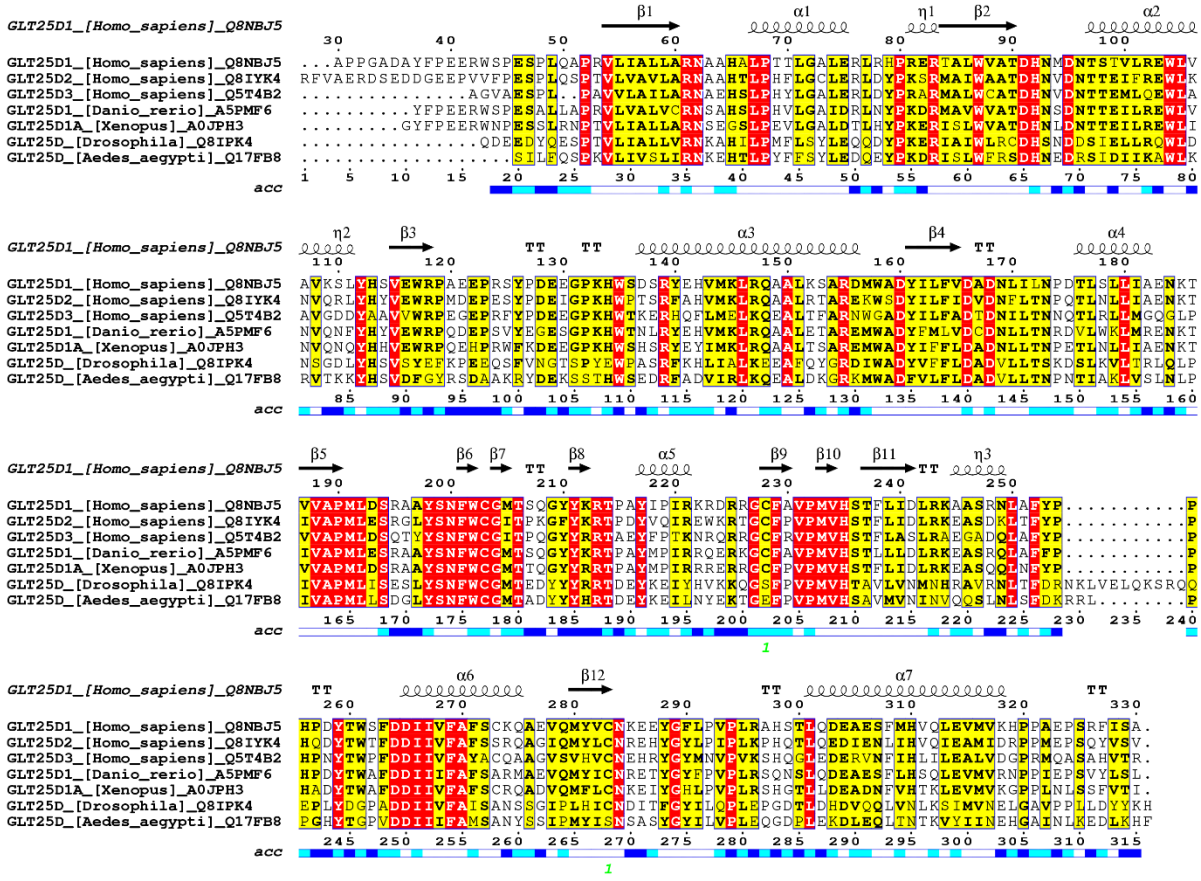
Supplementary Figure 3. GLT25D1/COLGALT1 is a Mn^{2+}-dependent galactosyltransferase. (a) Endpoint luminescence-based assays were used to evaluate the galactosyltransferase enzymatic activity in presence of different metal ions. Control experiments were performed without adding GLT25D1/COLGALT1. (b) Steady-state enzyme kinetics analysis was performed using Mn^{2+} or Mg^{2+} and UDP-Gal donor substrate, highlighting a strong preference for Mn^{2+} or Mg^{2+} as catalytic metal ion cofactor. Quantitation of K_M and k_{cat} was carried out using nonlinear fit through Michaelis-Menten equation in *Graphpad Prism*[1]. Results are shown in Supplementary Table 5. (c) Endpoint luminescence-based assays were used to evaluate the galactosyltransferase enzymatic activity in presence of different donor substrates and their analogs. Control experiments were performed without adding GLT25D1/COLGALT1. Uncoupled measurements were carried out in absence of acceptor (i.e., gelatin) substrate. In all figure panels, error bars represent standard deviations from average of triplicate independent experiments.



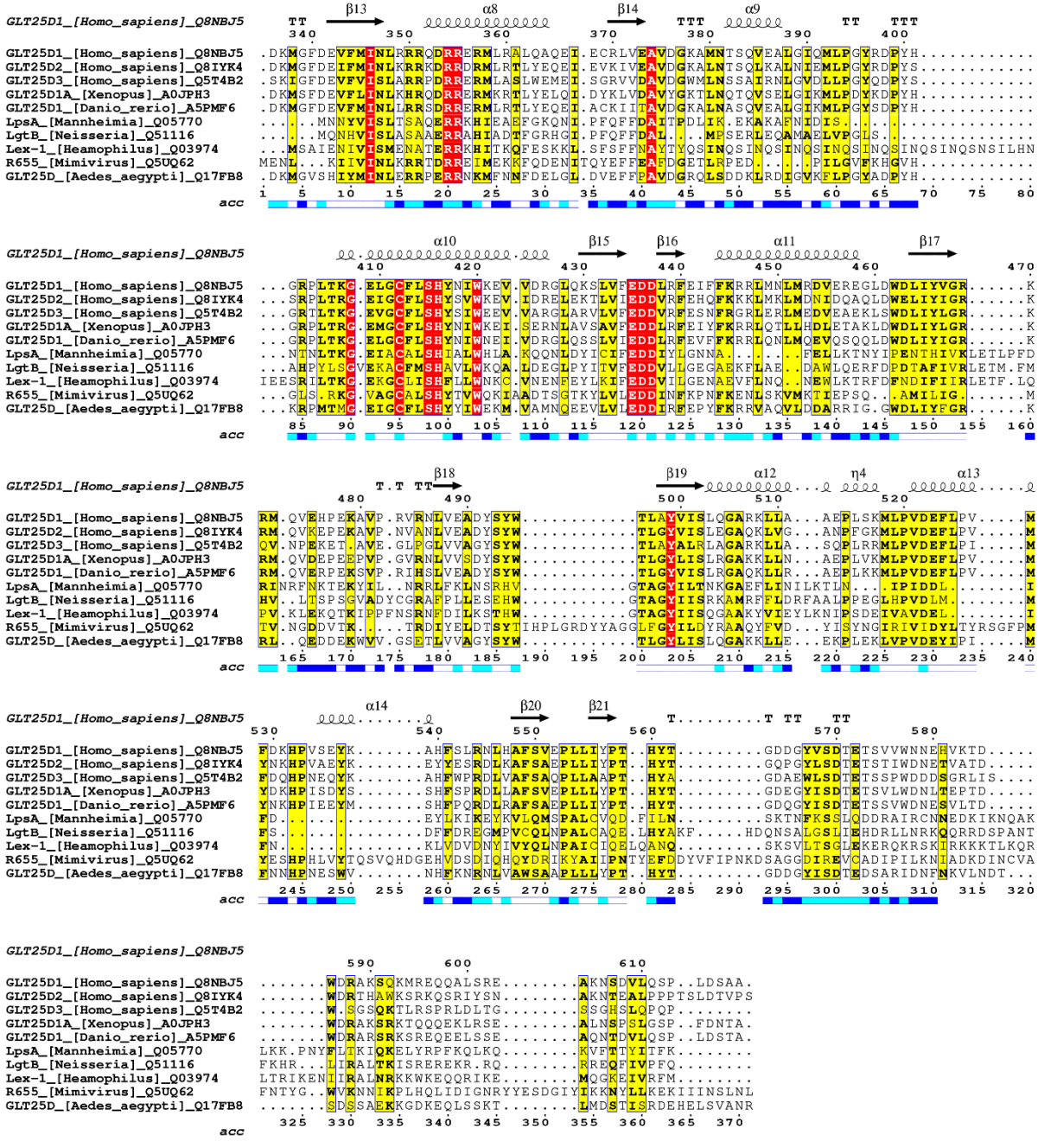
Supplementary Figure 4. Quality of the experimental electron density for the crystal structure of full-length human GLT25D1/COLGALT1. Shown are snapshots of the refined experimental electron density ($2F_o$-F_c, contour level 1 σ) used to generate the GLT25D1/COLGALT1 structural model using the Mn^{2+}, UDP-α-Gal co-crystallized model, covering the GT1 domains (a,d), the GT2 domains (b,e), and the linker regions (c,f) for each of the two copies of the enzyme found in the asymmetric unit.



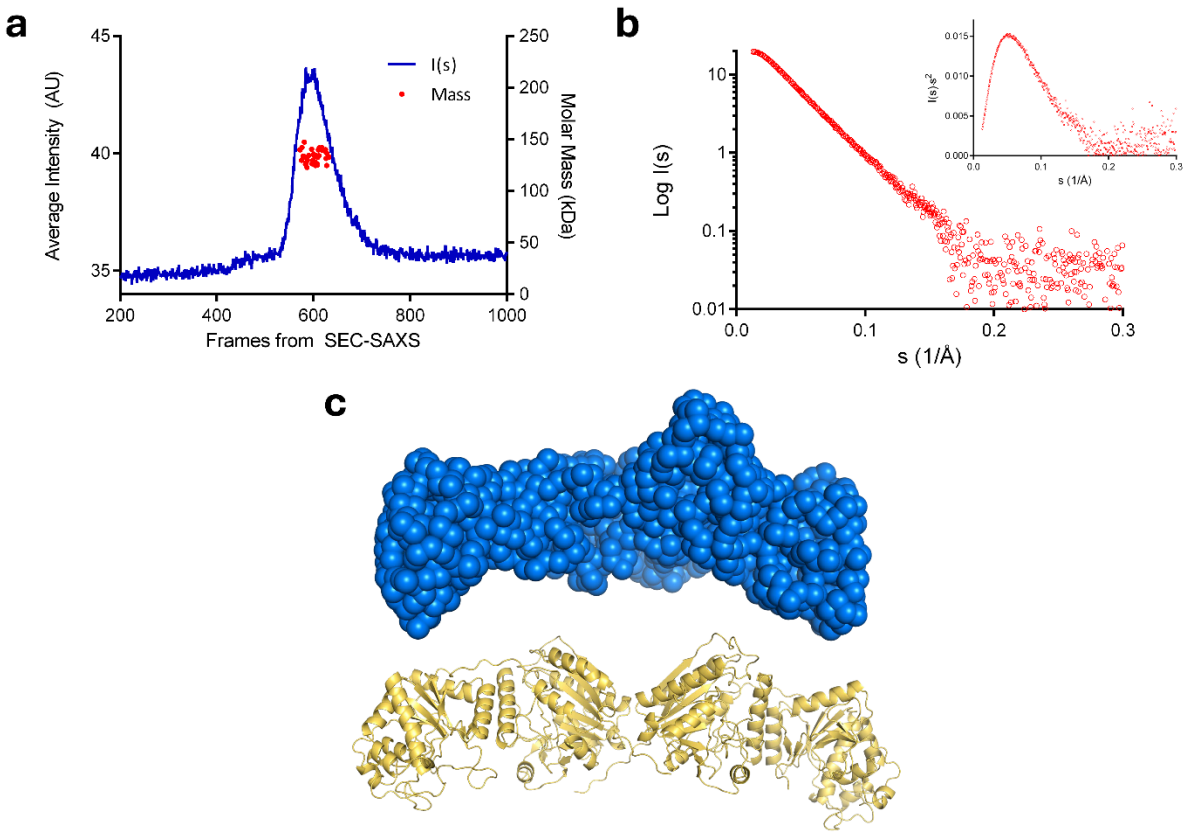
Supplementary Figure 5. Topological details of GLT25D1/COLGALT1 GT1 and GT2 domains. (a) Topology diagrams for the GT1 (teal) and GT2 (gold) domains, respectively, drawn using TOPDRAW [2]; (b) Superposition of structural homologs to the GT1 domain of GLT25D1/COLGALT1 (shown as thick black ribbon), as identified using DALI with Z-score higher than 15. The table inset indicates the PDB ID of each entry, as well as the associated DALI Z-scores and RMSD values. (c) The closest structural homolog of GLT25D1/COLGALT1 GT1 domain is the AC domain of human LH3/PLOD3. For reference, the image shown on the right highlights the position of the AC domain within the LH3/PLOD3 polypeptide (PDB ID 6FXR). The superposition shows the excellent match between the two molecular architectures (shown in teal for GLT25D1/COLGALT1 and orange for LH3/PLOD3, respectively) and the key differences (shown with a black arrow) in the region proximate to the cofactor and donor substrate binding site identified in GLT25D1/COLGALT1. (d) Superposition of structural homologs to the GT2 domain of GLT25D1/COLGALT1, as identified using DALI with Z-score higher than 10. The table inset indicates the PDB ID of each entry, as well as the associated DALI Z-scores and RMSD values. (e) Superposition of the GT2 domain of GLT25D1/COLGALT1 (magenta) with LH3/PLOD3 GT domain (blue) shows limited similarity with the N-terminal GT domain of the multifunctional enzyme, with DALI Z-score lower than 10 and RMSD of 4.7 Å.



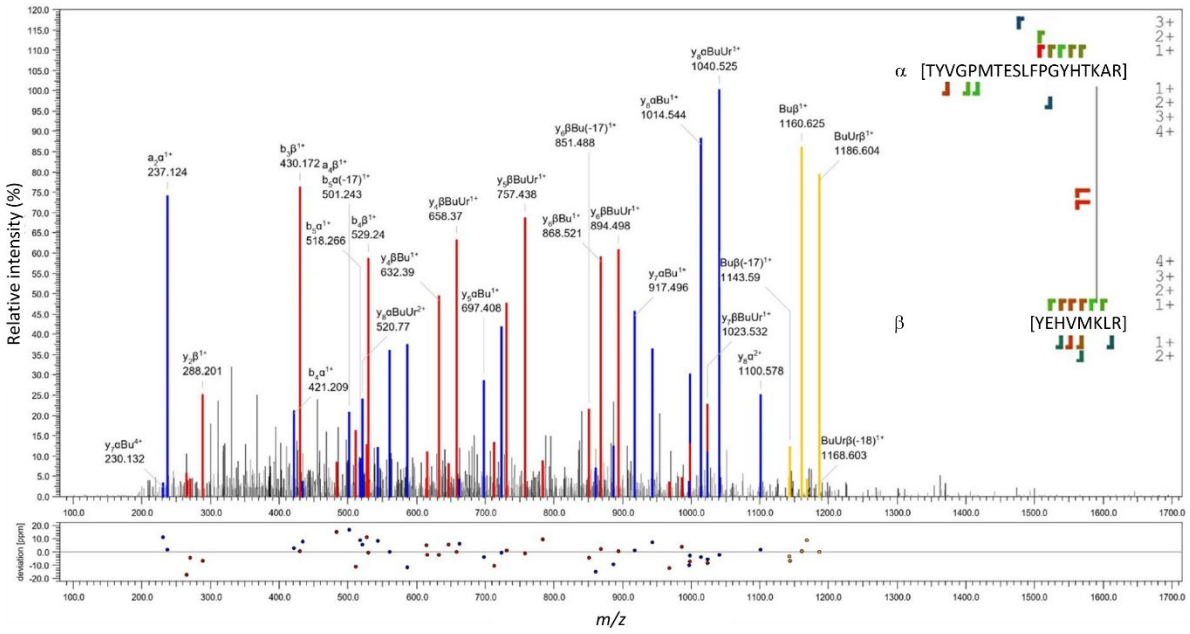
Supplementary Figure 6. Alignment of selected amino acid sequences from different species obtained from a BLAST search against the GLT25D1/COLGALT1 GT1 domain. The alignment has been generated using EBI MUSCLE[3] and drawn using ESPRIPT[4]. The "*consensus*" line computes the sequence consensus according to the result of the multiple sequence alignment (global score 0.7, differential score 0.5). The "*acc*" line indicates the solvent accessibility of each residue based on evaluation of the experimental structure of human GLT25D1/COLGALT1 (white, non-accessible residues; blue, accessible residues).



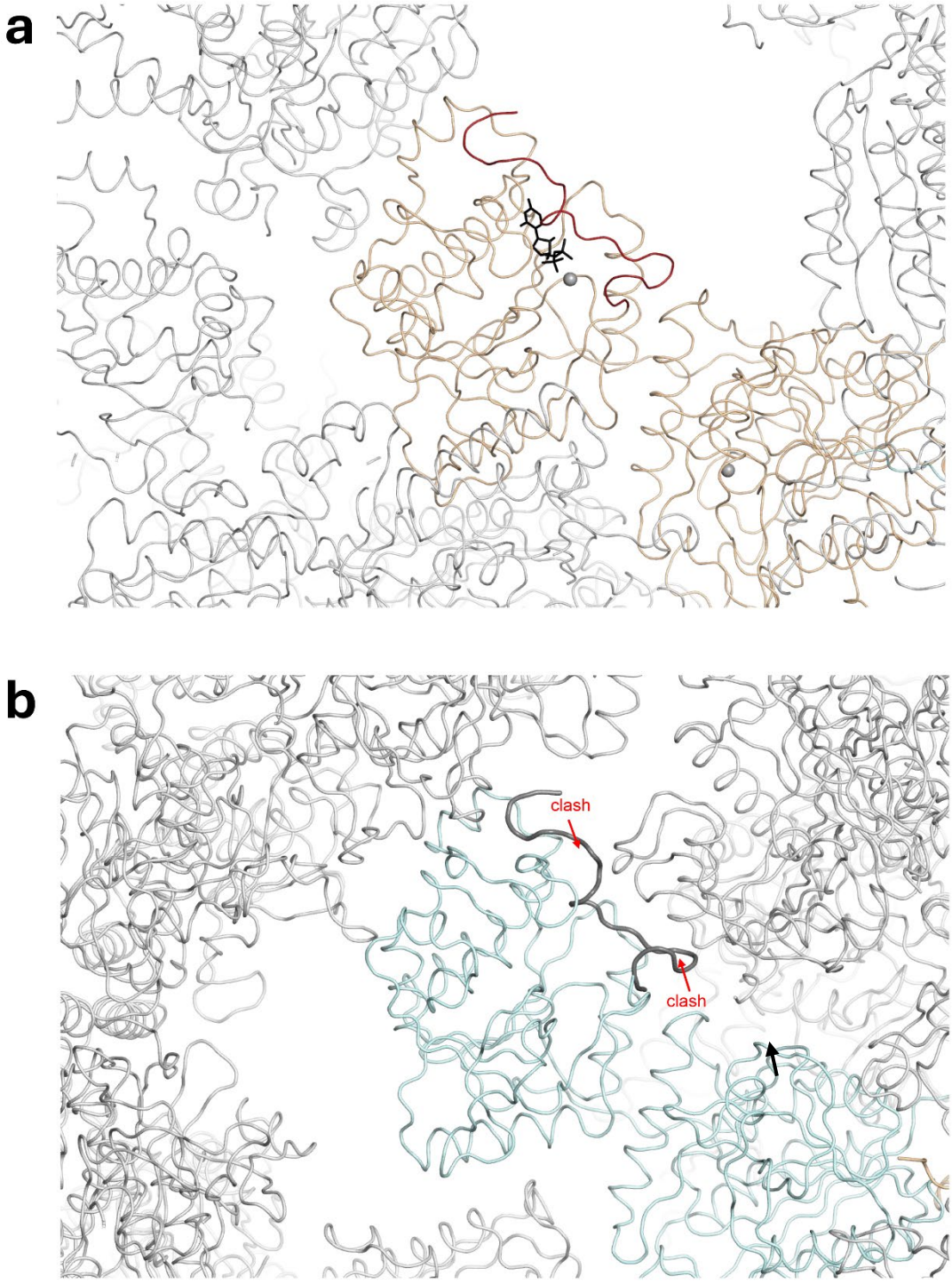
Supplementary Figure 7. Alignment of selected amino acid sequences from different species obtained from a BLAST search against the GLT25D1/COLGALT1 GT2 domain. The alignment has been generated using EBI MUSCLE[3] and drawn using ESPRIPT[4]. The "*consensus*" line computes the sequence consensus according to the result of the multiple sequence alignment (global score 0.7, differential score 0.5). The "*acc*" line indicates the solvent accessibility of each residue based on evaluation of the experimental structure of human GLT25D1/COLGALT1 (white, non-accessible residues; blue, accessible residues).



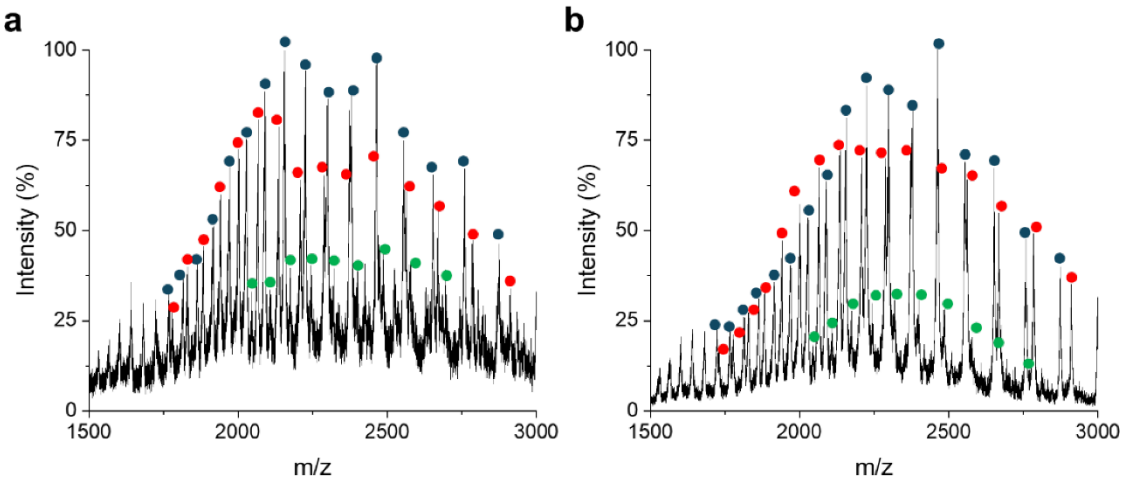
Supplementary Figure 8. SEC-SAXS analysis of GLT25D1/COLGALT1 in solution. (a) SEC-SAXS chromatogram and molecular weight analysis of peak fractions derived from CHROMIXS analysis[5]. (b) SAXS data obtained from the analysis of the peak frames of the SAXS chromatogram shown in (a) (red circles, with Kratky transformation in the inset). (c) *Ab-initio* sphere models based on solution SEC-SAXS data match the observed dimeric arrangement found in the experimental crystal structure of GLT25D1/COLGALT1. Shown is the comparison of the size and shape of an *ab initio* envelope computed from SEC-SAXS data (blue spheres) using *GASBOR*[6] with the crystal structure of GLT25D1/COLGALT1 (gold cartoon).



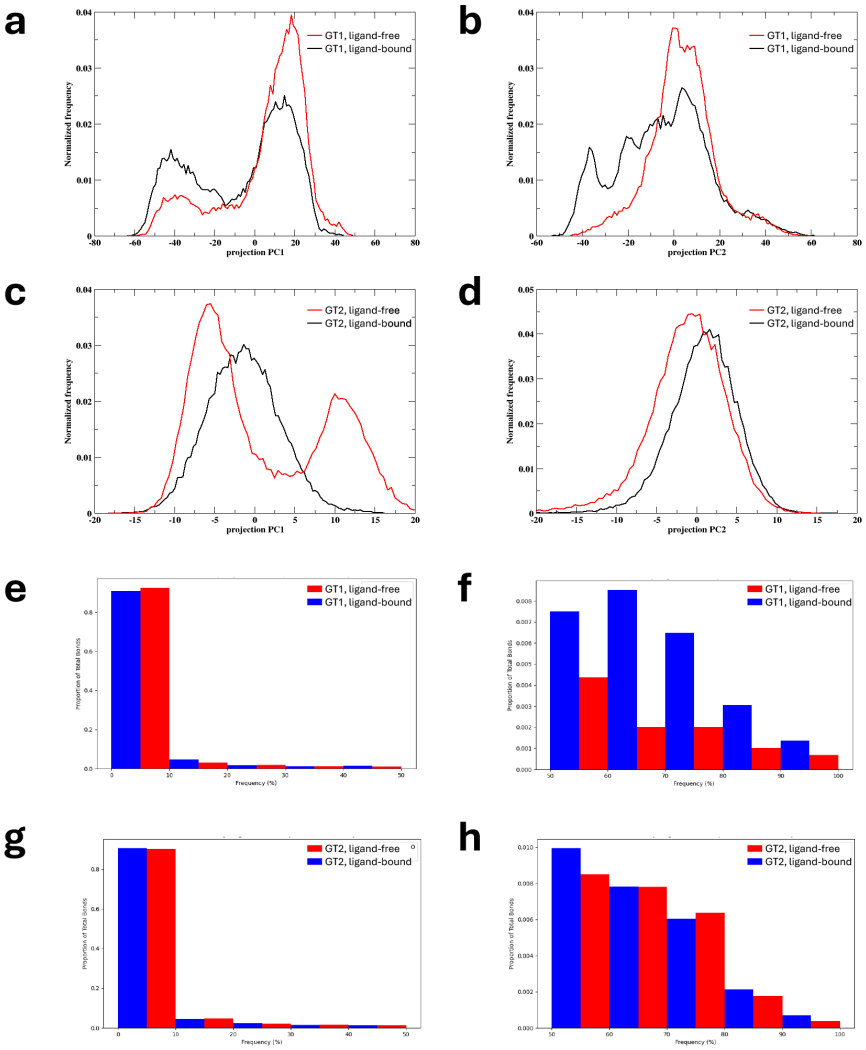
Supplementary Figure 9. Results of the XL-MS analysis of GLT25D1/COLGALT1 and LH3/PLOD3. Product ion mass spectrum of the cross-linked peptides (α) [TYVGPMTESLFPGYHT^{645}KAR] of LH3/PLOD3 and (β) [YEHVM^{145}KLR] of GLT25D1/COLGALT2. In the spectrum, the characteristic DSBU linker-derived fragment ions are highlighted in yellow, while fragment ions originating from the α peptide and β peptide are shown in blue and red, respectively. The precursor ion was detected with a mass accuracy of 1.0 ppm. The bottom panel reports the mass accuracy of fragment ions. The cross-linked species $[M+5H]^{5+}$ at *m/z* 685.9492 was subjected to CID-MS/MS, and fragment ions were annotated using the MeroX software[7]. The cross-link fragmentation profile displayed on the right shows the identified fragment ions with their respective positions in the peptide, as well as the charge. The relative intensity of the identified ion is color-coded from 0% (blue) to 100% (red). Data were obtained from file LH3_GLT_8_GB4_1_2907.d of the PRIDE dataset accompanying this paper, scan ID 25363, with a retention time of 5593.19 s (93.2 min ± 0.4 min) and an inverse ion mobility value of $1/K_0$ = 0.978.



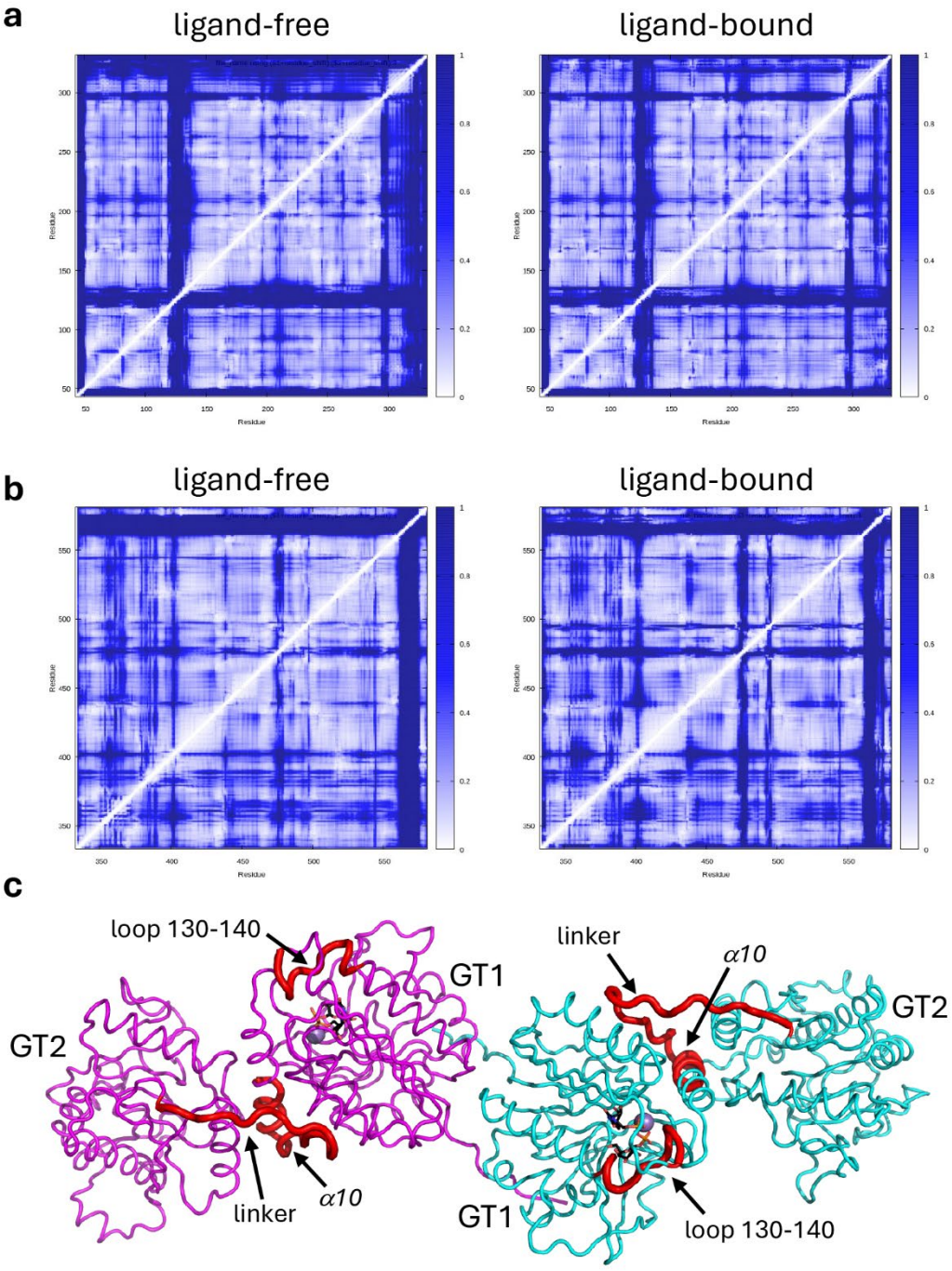
Supplementary Figure 10. Overview of the crystal packing observed in the crystal structures of GLT25D1/COLGALT1. The representations highlight the different contacts surrounding the GT2 domains in the two copies of the enzyme found in the asymmetric unit of the enzyme structure obtained with excess Mn^{2+} and UDP-α-Gal. The conformation adopted by the 560-570 loop in the substrate-bound structure, compatible with the packing of monomer A shown is panel (a), is not compatible with the crystal packing of the second monomer (b). In B, the loop from superposed monomer A is shown as thick grey ribbon to highlight the packing clashes.



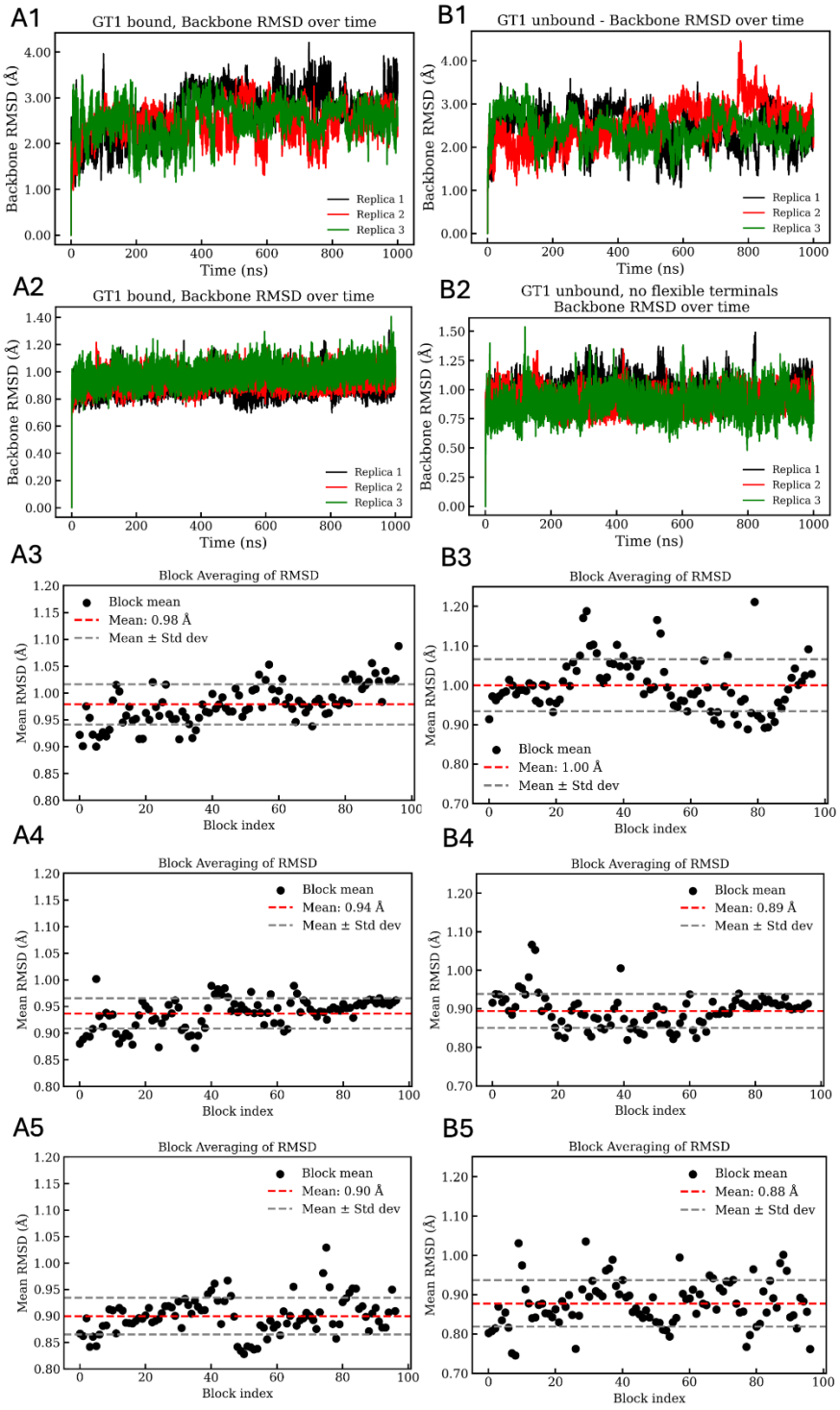
Supplementary Figure 11: Native-MS analysis. Native-MS spectra of wild-type (a) and Trp158Arg (b) GLT25D1/COLGALT1 under partially denaturing conditions (20 µM protein concentration in 50 mM ammonium bicarbonate, 40% acetonitrile). A charge-state distribution of monomeric, partially-unfolded conformation was detected (blue dots), with experimental molecular weights (68,969 ± 8 Da in panel a, 68,942 ± 2 Da in panel b) in close agreement with theoretical ones (68,972.7 Da and 68,942.6 Da, respectively). Putative GLT25D1/COLGALT1 deleted forms (red dots) of 64,049 ± 5 Da (a) and 64,019 ± 10 Da (b) were also present, confirming the doublet bands detected in SDS-PAGE analysis (Supplementary Figure 1). A minor molecular population, ascribable to protein-ligand complex, was also identified (greed dots), with molecular weigths of 69,575 ± 7 Da (a) and 69,550 ± 4 Da (b), compatible with protein interaction with UDP-α-Gal and Ca^{2+}.



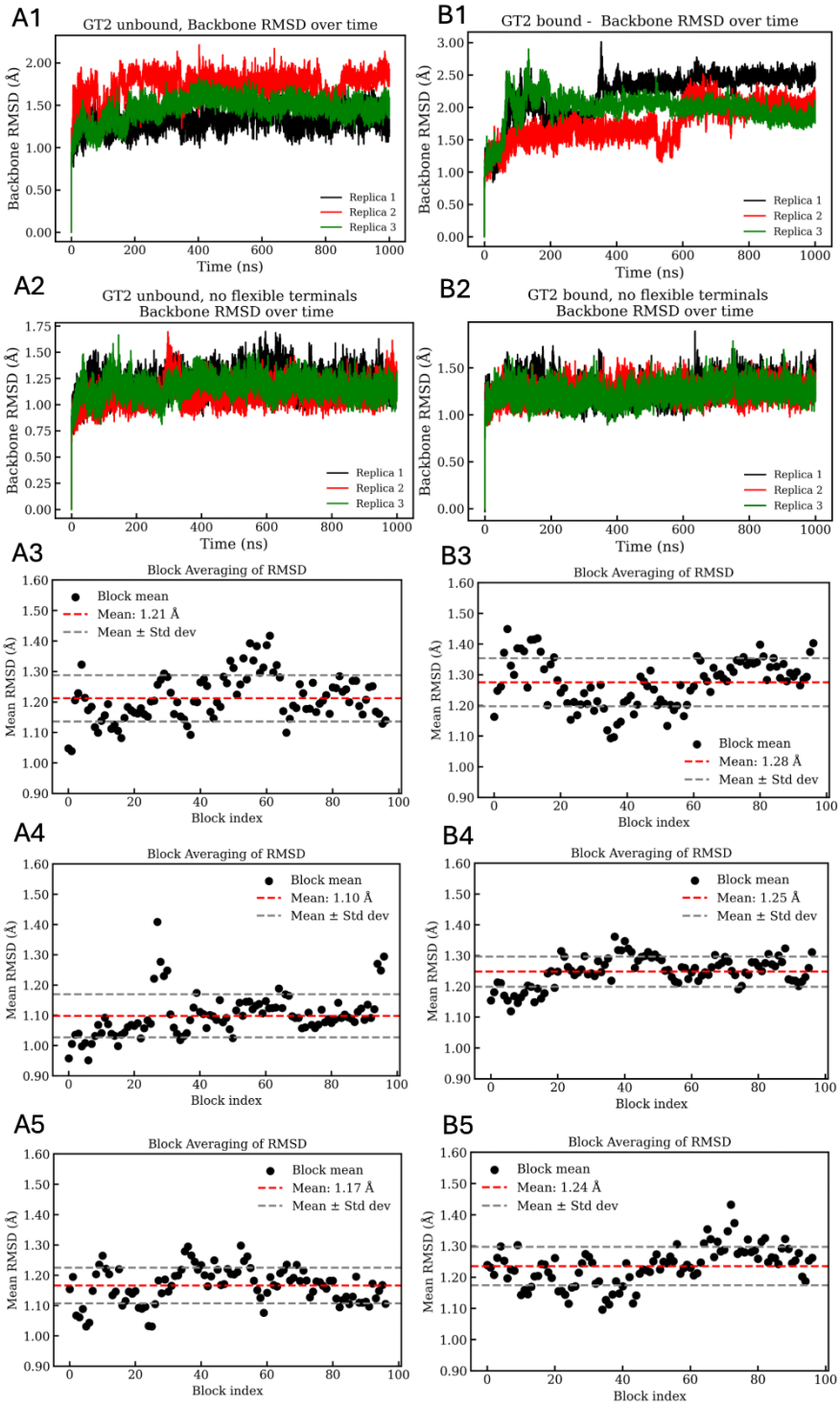
Supplementary Figure 12. Analysis of the MD simulations data. (a,b) Principal Component Analysis (PCA) of the GT1 domain. Normalized frequency distributions of the projections of the molecular dynamics trajectories of the GT1 domain in the unbound (black) and ligand-bound (red) states onto the first two principal components PC1 (a) and PC2 (b), which together capture 79% of the total variance in protein motion. (c,d) Principal Component Analysis (PCA) of the GT2 domain. Normalized frequency distributions of the projections of the molecular dynamics trajectories of the GT2 domain in the unbound (black) and ligand-bound (red) states onto the first two principal components PC1 (c) and PC2 (d), which together capture 84% of the total variance in protein motion. (e,f) Normalized histograms of hydrogen bond persistence in the GT1 domain for the ligand-bound (blue) and unbound (red) states. The normalized frequency persistences were calculated and grouped into 10% intervals; panel (e) represents the normalized frequency of hydrogen bonds persisting for 0–50% of the simulation time, while panel (f) represents those persisting for 50–100%. Frequencies were normalized to account for the total number of hydrogen bonds detected, allowing for a direct comparison between the two states. (g,h) Normalized histograms of hydrogen bond persistence in the GT2 domain for the ligand-bound (blue) and unbound (red) states. The normalized frequency persistences were calculated and grouped into 10% intervals; panel (g) represents the normalized frequency of hydrogen bonds persisting for 0–50% of the simulation time, while panel (h) represents those persisting for 50–100%. Frequencies were normalized to account for the total number of hydrogen bonds detected, allowing for a direct comparison between the two states.



Supplementary Figure 13. MD simulations highlight enhanced flexibility in specific areas of the GT1 domain. (a-b) Characterization of the internal dynamics and flexibility of the simulated domains of GLT25D1/COLGALT1 in terms of residue-pair fluctuations. (a) Matrix of residue-pair distance fluctuations (DF) for the GT1 domain in the ligand-free (left) and ligand-bound (right) states. (b) Matrices of residue-pair DFs for the GT2 domain in the ligand-free (left) and ligand-bound (right) states. The matrices are color-coded from white to dark blue, with darker shades representing greater distance fluctuations, less coordinated motion between residues and higher flexibility. . (c) Localization of the most prominent protein segments showing increased flexibility and less restricted to a well-defined conformation in molecular dynamics simulations upon removal of metal ions and donor substrates from the GLT25D1/COLGALT1 GT1 domain.



Supplementary Figure 14 - Equilibration analysis of MD simulations of GT1 in unbound and bound states. (A1) Backbone RMSD of GT1 in the unbound state across the three independent simulation replicates, calculated over the production phase of the simulations. (A2) Backbone RMSD of GT1 after excluding the flexible terminal regions (first and last 10 residues), also computed over the production phase. (A3–A5) Block-averaging analysis of the backbone RMSD for the unbound GT1 domain, in replicate 1, replicate 2, and replicate 3, respectively, assessing statistical convergence. (B1–B5) Equivalent analyses for GT1 in the bound state, with (B1) showing the backbone RMSD of the full domain, (B2) the backbone RMSD excluding the flexible terminal regions, and (B3–B5) the block-averaging analysis.



Supplementary Figure 15 MD equilibration - Equilibration analysis of MD simulations of GT2 in unbound and bound states. (A1) Backbone RMSD of GT2 in the unbound state across the three independent simulation replicates, calculated over the production phase of the simulations. (A2) Backbone RMSD of GT2 after excluding the flexible terminal regions (first 10 and last 18 residues), also calculated over the production phase. (A3–A5) Block-averaging analysis of the RMSD for the unbound GT2 domain in replicate 1, replicate 2, and replicate 3, respectively, assessing statistical convergence. (B1–B5) Equivalent analyses for GT2 in the bound state, with (B1) showing the backbone RMSD of the full domain, (B2) the backbone RMSD excluding the flexible terminal regions, and (B3–B5) the block-averaging analysis for replicate 1, replicate 2, and replicate 3, respectively.

SUPPLEMENTARY TABLES

Supplementary Table 1. List of oligonucleotides used to generate GLT25D1/COLGALT1 mutants

Site	Mutation	Direction	Sequence (5'→3')
GT1, dimer interface	Trp158Arg	Forward	ggctgattacatcctgtttgtagat
		Reverse	CTaccgagacctcagtcgtatggaacaat
GT1, cofactor binding site	Trp135Arg	Forward	cgtctgactcacgctacgagcatgtcat
		Reverse	Ggtgtttcgggccttcctcgtccgggt
	Asp166Ala	Forward	CAgcggacaacctgatcctcaaccc
		Reverse	ctacaaacaggatgtaatcagcccacatgt
	Asp265Ala	Forward	catcatcgtctttgccttctcctgcaag
		Reverse	Gcgtcaaaggaccaggtgtagtcagggt
	Ile266Gln	Forward	CAAatcgtctttgccttctcctgcaag
		Reverse	gtcgtcaaaggaccaggtgtagtcaggg
	Ile267Gln	Forward	CAAgtctttgccttctcctgcaagcagg
		Reverse	gatgtcgtcaaaggaccaggtgtagt
GT2, cofactor binding site	Arg351Ala	Forward	GCTcaggaccggcgggagcgcatgctgc
		Reverse	ccgcctcaggttaatcatgaagacctcgt
	Cys412Ser	Forward	Gcttcctgagccactacaacatctggaag
		Reverse	cttccagatgttgtagtggctcaggaagc
	Glu435Ala	Forward	ggatgacctgcgttttgagatcttcttc
		Reverse	Gcaaacacaagcgatttctgcagcccc
	Asp437Ala	Forward	Ccctgcgttttgagatcttcttcaagagac
		Reverse	catcctcaaacacaagcgatttctgcag
	Trp495Ala	Forward	gaccctggcctacgtgatctccctgcaag
		Reverse	GCgtaggaatagtcagcttcgactaggtt
	Asp522Ala	Forward	cgagttcctgcccgtcatgttcgacaaac
		Reverse	Gctactggaagcattttggagagcggc
	Asp570Ala	Forward	CTaccgagacctcagtcgtatggaacaat
		Reverse	cactcacatagccatcgtctcctgtgta

Supplementary Table 2. X-ray diffraction data collection parameters and statistics

[a] Values in parentheses are for reflections in the highest resolution shell.

	GLT25D1/COLGALT1 (Native)	GLT25D1/COLGALT1 (co-crystallized with Mn^{2+}, UDP-α-Gal)	GLT25D1/COLGALT1 (Hg^{2+} Soak)
Data Collection[a]			
X-ray source	ESRF ID23-1	SLS X06SA	ESRF ID23-1
Processing programs	*XDS*, *AIMLESS*	*XDS*, *AIMLESS*	*XDS*, *AIMLESS*, *SHELX, HKL2MAP*
Space group	I23	I23	I23
Cell parameters	*a* = 221.9 Å *a* = 90.0° *b* = 221.9 Å *b* = 90.0° *c* = 221.9 Å *g* = 90.0°	*a* = 220.8 Å *a* = 90.0° *b* = 220.8 Å *b* = 90.0° *c* = 220.8 Å *g* = 90.0°	*a* = 220.2 Å *a* = 90.0° *b* = 220.2 Å *b* = 90.0° *c* = 220.2 Å *g* = 90.0°
Wavelength (Å)	0.8856	0.9999	0.7749
Resolution range (Å)	49.62-3.00 (3.13-3.00)	49.38-2.70 (2.79-2.70)	49.24-2.80 (2.91-2.80)
Total reflections	123628 (15938)	425956 (23149)	1013021 (105836)
Unique reflections	35735 (4385)	49018 (4444)	43652 (4551)
CC1/2[b]	0.992 (0.402)	0.998 (0.395)	0.998 (0.599)
Redundancy	3.5 (3.6)	8.7 (5.2)	23.2 (23.3)
Mean I/σ(I)	6.2 (1.0)	13.5 (1.1)	13.2 (0.9)
Completeness (%)	98.4 (99.9)	99.7 (97.9)	100.0 (100.0)
R_{sym}[b]	0.130 (1.047)	0.110 (1.251)	0.169 (3.609)
R_{pim}[c]	0.101 (0.811)	0.058 (0.920)	0.051 (1.093)
Anomalous completeness (%)			100.0 (100.0
Anomalous multiplicity			11.7 (11.7)
Anomalous CC1/2			0.488
SAD resolution limit (Å)[d]			4.4

[b] $R_{sym} = [\Sigma_{hkl}\Sigma_j | I_{hkl,j} - \langle I_{hkl} \rangle |] / [\Sigma_{hkl}\Sigma_j I_{hkl,j}]$, where I is the observed intensity for a reflection and <I> is the average intensity obtained from multiple observations of symmetry-related reflections.

[c] $R_{pim} = [\Sigma_{hkl} (1/(n-1))^{1/2} \Sigma_j | I_{hkl,j} - \langle I_{hkl} \rangle |] / [\Sigma_{hkl}\Sigma_j I_{hkl,j}]$ where I is the observed intensity for a reflection and <I> is the average intensity obtained from multiple observations of symmetry-related reflections.

[d] the resolution limit for SAD phasing has been selected by evaluating the resolution where *d''/s* falls below 1.2 using the *SHELXC* module of *HKL2MAP*.

Supplementary Table 3. X-ray crystallographic refinement statistics

	GLT25D1/COLGALT1 (Native)	GLT25D1/COLGALT1 (co-crystallized with Mn^{2+}, UDP-a-Gal)	GLT25D1/COLGALT1 (Hg^{2+} Soak)
Refinement			
R_{work}/R_{free}[c]	0.1970/0.2417	0.1846/0.2216	0.2024/0.2390
Number of atoms:	8633	8923	8621
Protein	8519	8681	8503
Ligands	105	106	114
Solvent	9	136	4
Average B-factor $(Å)^2$	92.07	74.93	107.55
Protein	92.22	75.27	107.58
Ligands	81.40	64.46	105.82
Solvent	64.47	54.66	85.53
Structure quality			
RMS bond lengths (Å)	0.003	0.002	0.003
RMS bond angles (°)	0.63	0.52	0.55
Ramachandran stats			
Favored (%)	96.03	96.76	95.74
allowed (%)	3.97	3.14	4.26
outliers (%)	0.00	0.10	0.00
PDB ID	9EVK	9EVJ	9EVL

Supplementary Table 4. Details of mass photometry analysis

GLT25D1/COLGALT1 Sample	Shown in	Mass (kDa)	Intensity Counts	Peak mass fraction
wild type	Figure 2B	142.0 ± 8.8	2991	85%
wild type	Figure 3B	137.0 ± 10.4	4597	83%
Trp158Arg	Figure 3B	61.0 ± 7.6	5213	88%
wild type	Figure 6c	131.0 ± 11.4	3236	70%
wild-type, EDTA treated	Figure 6c	124.0 ± 17.2	6703	84%

Supplementary Table 5. Summary of the outcome of the luminescence-based indirect activity assays, and the HRMS direct assays to evaluate GLT25D1/COLGALT1 enzymatic activity. Values are reported as percentage of enzymatic activity observed for each variant compared to that observed using the wild-type enzyme. Values are expressed as average of minimum triplicate independent measurements, with associated standard deviations.

GLT25D1/COLGALT1 variant	Gal-T activity			
	Luminescence-based indirect assay			HRMS direct assay (% activity)
	K_M (mg mL^{-1})	k_{cat} (min^{-1})	k_{cat}/K_M (mL min^{-1} mg^{-1})	
wild-type	5.61 ± 1.71	67272 ± 6060	11991 ± 3811	100.0 ± 22.1
Trp135Arg	n.d.	n.d.	n.d.	n.d.
Trp158Arg	13.37 ± 4.45	73849 ± 10321	5523 ± 1994	97.2 ± 4.3
Ile267Gln	15.21 ± 7.29	82113 ± 17457	5398 ± 2831	93.3 ± 8.9
Arg315Ala	n.d.	n.d.	n.d.	n.d.
Cys412Ser	16.35 ± 9.20	1748 ± 450	107 ± 66	n.d.
Glu435Ala	n.d.	n.d.	n.d.	n.d.
Asp437Ala	n.d.	n.d.	n.d.	n.d.
Trp495Ala	n.d.	n.d.	n.d.	n.d.
Asp522Ala	n.d.	n.d.	n.d.	n.d.
Asp570Ala	n.d.	n.d.	n.d.	n.d.
wild-type using Mg^{2+}	7.23 ± 1.23	16829 ± 933	2327 ± 416	n.d.

Supplementary Table 6. SEC-SAXS data collection parameters and processing statistics

Data Collection	
Beamline	ESRF BM29
Beam energy (keV)	12.5
Sample-detector distance (m)	2.867
Exposure time (s)	1
Sample cell thickness (mm)	1
Temperature (°)	20
Final q range (nm^{-1})	0.01 – 4.00
Data Analysis	
Points used for Guinier analysis	10 - 33
Guinier qR_g limits	0.77 - 1.28
Guinier R_g (nm)	4.33 ± 0.02
I(0) (mm^{-1})	22.91 ± 0.08
D_{max} (nm)	13.97
MW estimation (Bayesian Inference) (kDa)	131 ± 10

Supplementary Table 7: MD systems setup

	GT1 unbound	GT1 bound	GT2 unbound	GT2 bound
Simulation box dimensions	85.39 Å x 85.39 Å x 85.39 Å 109.47° x 109.47° x 109.47°	83.44 Å x 83.44 Å x 83.44 Å 109.47° x 109.47° x 109.47°	71.80 Å x 71.80 Å x 71.80 Å 109.47° x 109.47° x 109.47°	71.92 Å x 71.92 Å x 71.92 Å 109.47° x 109.47° x 109.47°
Total number of atoms	42292	39396	24407	24542
Total number of water molecules	12524	11541	6756	6789
Salt concentration	150 mM	150 mM	150 mM	150 mM

Supplementary Table 8. Molecular dynamics simulations checklist

Reliability and reproducibility checklist for molecular dynamics simulations ***All boxes must be marked YES by acceptance unless "Response not needed if No".**	**Yes**	**No**	**Response** **(Please state where this information can be found in the text)**
1. Convergence of simulations and analysis			
1a. Is an evaluation presented in the text to show that the property being measured has equilibrated in the simulations (*e.g.* time-course analysis)?	☒	☐	Methods, *"Molecular dynamics simulations";* Supplementary Methods; Main text, "*MD simulations supports distinct roles for GT1 and GT2 domains".*
1b. Then, is it described in the text how simulations are split into equilibration and production runs and how much data were analyzed from production runs?	☒	☐	Methods, *"Molecular dynamics simulations";* Supplementary Methods.
1c. Are there at least 3 simulations per simulation condition with statistical analysis?	☒	☐	Methods, *"Molecular dynamics simulations";* Supplementary Methods.
1d. Is evidence provided in the text that the simulation results presented are independent of initial configuration?	☒	☐	Methods, *"Molecular dynamics simulations";* Supplementary Methods; Main text, "*MD simulations supports distinct roles for GT1 and GT2 domains".*
2. Connection to experiments			
2a. Are calculations provided that can connect to experiments (*e.g.* loss or gain in function from mutagenesis, binding assays, NMR chemical shifts, J-couplings, SAXS curves, interaction distances or FRET distances, structure factors, diffusion coefficients, bulk modulus and other mechanical properties, *etc*.)?	☒	☐	Main text, "*MD simulations supports distinct roles for GT1 and GT2 domains".* All analyses aim to assess whether the biochemical data are consistent with the simulations, providing a perspective to support experimental findings.
3. Method choice			
3a. Do simulations contain membranes, membrane proteins, intrinsically disordered proteins, glycans, nucleic acids, polymers, or cryptic ligand binding?	☐	☒	Response not needed if **No**
3b. Is it described in the text whether the accuracy of the chosen model(s) is sufficient to address the question(s) under investigation (e.g. all-atom vs. coarse-grained models, fixed charge vs. polarizable force fields, implicit vs. explicit solvent or membrane, force field and water model, etc.)?	☐	☐	Main text, "*MD simulations supports distinct roles for GT1 and GT2 domains".*
3c. Is the timescale of the event(s) under investigation beyond the brute-force MD simulation timescale in this study that enhanced sampling methods are needed?	☐	☒	Main text, "*MD simulations supports distinct roles for GT1 and GT2 domains".*
If **YES**, are the parameters and convergence criteria for the enhanced sampling method clearly stated?	☐	☐	
If **NO,** is the evidence provided in the text?	☒	☐	Main text, "*MD simulations supports distinct roles for GT1 and GT2 domains".*
4. Code and reproducibility			
4a. Is a table provided describing the system setup that includes simulation box dimensions, total number of atoms, total number of water molecules, salt concentration, lipid composition (number of molecules and type)?	☒	☐	Methods, *"Molecular dynamics simulations";* Supplementary Methods.

4b. Is it described in the text what simulation and analysis software and which versions are used?	☒	☐	Methods, *"Molecular dynamics simulations";* Supplementary Methods.
4c. Are other parameters for the system setup described in the text, such as protonation state, type of structural restraints if applied, nonbonded cutoff, thermostat and barostat, etc.?	☒	☐	Methods, *"Molecular dynamics simulations";* Supplementary Methods.
4d. Are initial coordinate and simulation input files and a coordinate file of the final output provided as supplementary files or in a public repository?	☒	☐	Data availability statement. The initial coordinate files, simulation input files, and final output coordinate files are available in the public repository Zenodo at the following DOI: 10.5281/zenodo.15020231
4e. Is there custom code or custom force field parameters?	☐	☒	Response not needed if **No**
If **YES**, are they provided as supplementary files or in a public repository?	☐	☐	

Supplementary Methods

System preparation for molecular dynamics simulations

The molecular systems were prepared based on the crystal structures of the GT1 and GT2 domains of the enzyme. Both domains contain Mn^{2+} and UDP-Gal as cofactors in their crystallographic structures. For the C-terminal domain, only UDP was retained, as electron density for the Gal moiety was insufficient to accurately model it.

Five different systems were prepared for simulation: (1) GT1 domain with cofactors (Mn^{2+} and UDP-Gal); (2) GT1 domain without cofactors (cofactors removed from the structure); (3) GT2 domain with cofactors (Mn^{2+} and UDP); (4) GT2 domain without cofactors (cofactors removed from the structure). For simulations in the absence of cofactors, the Mn^{2+} and UDP-Gal (or UDP in the case of the GT2 domain) were removed from the crystallographic complex structures to mimic the apo form. The parameters and charges for UDP and UDP-Gal were generated using *Antechambe*r tool within the *AmberTools* suite (version 24)[8,9], applying the General AMBER Force Field (GAFF)[10], which provides broad compatibility with organic molecules. Atom types and charges were assigned using the the AM1-BCC2 charge method[11], ensuring compatibility with the AMBER ff19SB[12] force field used for the protein. For Mn^{2+}, the parameters developed by Bradbook and co-workers were applied[13]. These parameters were then incorporated into topology file to allow complete system simulations. The *tleap* utility in the *AmberTools* suite (version 24)[9] was used to add hydrogens to all systems, with the protonation states of titratable side chains of Asp, Glu, Arg, Lys, Tyr, His and Cys residues predicted using the H++[14] program at pH 7.0. In the GT1 domain, cysteine residues were evaluated, and a disulfide bond was added between CYS 228 ans CYS 283. Histidine residues 92, 142 and 319 were protonated at both Nε2 and Nδ1, while remaining histidine residues were protonated at Nε2. In the GT2 domain, no disulfide bonds were present. Histidine residues 401, 476 and 547 were protonated at both Nε2 and Nδ1, histidine 540 was protonated at Nδ1, while all other histidine residues were protonated at Nε2. The glutamate residue 523 was predicted in its neutral form. The systems were solvated in a truncated octahedral box with *TIP3P*[15] water molecules ensuring a minimum distance of 10 Å between the protein and the box edges and neutralized by adding Na^+ or Cl^- counter ions modeled using the parameters developed the by Joung and Cheatham[16] for the use with *TIP3P*[15] water. Additional Na^+ Cl^- ions were added to achieve an ionic concentration of 150 mM, distributed randomly throughout the solvent box to mimic physiological ionic conditions. A detailed summary of the system setup is provided in Supplementary Table 7.

Molecular dynamics simulations

All MD simulations in this study were performed using the *Amber* software package (version 20)[8,9], with the ff19SB force field[12]. The *sander* utility was used for the initial minimization and solvent equilibration phases, while GPU-accelerated *pmemd.cuda* utility was employed for all other stages of the simulations. Each system was simulated in three independent MD replicates (different random seeds to assign initial velocities) (Supplementary Fig. 14-15). Each replicate consisted of a 600-step minimization, a 2.069 ns preproduction time, and a production time of 1 μs. Before molecular dynamics production, all solvated systems underwent two sequential energy minimization steps to relieve steric clashes and unfavorable interactions. In the first minimization step, positional restraints (force constant of 5 kcal/mol/Å²) were applied to all heavy atoms, allowing only hydrogen atoms to move freely. The minimization was performed for a total of 300 steps, with 150 steps of steepest descent followed by 150 steps of conjugate gradient minimization. In the second minimization stage, the entire system was minimized without any positional restraints, also for 300 steps (150 steepest descent steps followed by 150 conjugate gradient steps). Both minimizations were conducted with a non-bonded interaction cutoff of 8.0 Å. Following energy minimization, the systems were equilibrated in the constant Number of particles (N), Volume (V), and Temperature (T), NVT, ensemble using a 9 ps simulated annealing protocol to optimize solvent packing around the solute. To this end, all solute atoms were restrained with a force constant of 10 kcal/mol/Å², allowing only the solvent molecules to equilibrate. The simulated annealing protocol consisted of three phases at different temperatires: heating phase (0–3 ps), the system was gradually heated from 25 K to 400 K; stable phase (3–6 ps), the

temperature was maintained at 400 K to allow for solvent relaxation; cooling phase (6–9 ps), the system was slowly cooled from 400 K to 25 K, facilitating optimal solvent packing around the solute. Temperature control was managed using the Berendsen thermostat[17], by adjusting the temperature coupling parameter: during heating and equilibration phases, a tight coupling with τ = 0.2 ps was applied to ensure stable temperature control, while during the cooling phase, a looser coupling was used, with τ = 2.0 ps initially, decreasing to τ = 1.0 ps for a smooth transition to lower temperatures. All equilibration steps were performed with a non-bonded interaction cutoff of 8.0 Å in periodic boundary conditions, and a 1 fs time step. The systems then were gradually heated from 25 to 300 K over a 20 ps molecular dynamics heating step in the NVT ensemble with periodic boundary conditions and a non-bonded interaction cutoff of 8.0 Å . Positional restraints (5 kcal/mol/Å²) were applied to the Cα atoms of the solute to maintain the overall structure. Temperature control was achieved with the Langevin thermostat, applying a collision frequency of 0.75 ps^{-1}. The SHAKE algorithm was applied to constrain all bonds involving hydrogen atoms, allowing for a 2 fs time step[18].

Following initial heating, the system underwent 2.04 ns of equilibration in the the constant Number of particles (N), Pressure (P), and Temperature (T) NPT ensemble (p = 1 bar) with periodic boundary conditions. The Berendsen barostat[17] was used for pressure control, and the temperature was maintained at 300 K with a Langevin thermostat[19] (collision frequency = 1.0 ps^{-1}). Non-bonded interactions were calculated with a cutoff of 8.0 Å, and SHAKE[18] constraints were applied to hydrogen bonds to allow a 2 fs time step. During equilibration, positional restraints on the Cα atoms were progressively reduced and then removed as follows: during the first 20 ps, restraints on Cα atoms were applied with a force constant of 3.75 kcal/mol/Å²; for the next 20 ps the restraint constant was reduced to 1.75 kcal/mol/Å² to allow for further structural relaxation; the final 2 ns, positional restraints were removed entirely, allowing all solute atoms to move freely. The production stage of MD simulations were conducted in the NPT ensemble with a 2 fs time step. Temperature was maintained at 300 K using a Langevin thermostat[19] (collision frequency of 1.0 ps^{-1}), while the pressure was held constant at 1 bar with Berendsen barostat[17] (relaxation time of 1.0 ps). Electrostatic interactions were computed directly within an 8.0 Å cutoff. Beyond this cutoff, Coulomb interactions were calculated with the particle mesh Ewald method[20]. Bond constraints for hydrogen-containing bonds were applied using the SETTLE algorithm for water molecules and SHAKE[18] for the solute.

To assess the equilibration of the generated trajectories, the time evolution of the RMSD of the protein backbone was analyzed for three independent simulation replicates of each system. The RMSD reached a plateau at different times depending on the system and replicate but did not exceed 25 ns in any case when considering the structured core of the domains, excluding the most flexible terminal regions (the first and last 10 residues for GT1, and the first 10 and last 18 residues for GT2). Therefore, the first 25 ns of the production phase were excluded from further analysis, ensuring that only the equilibrated portion of the trajectories was used.

To confirm equilibration, the decorrelation time of the RMSD in the remaining portion of each trajectory was determined by computing its autocorrelation function. The decorrelation time was determined as the first point where the autocorrelation function is below 1/e, yielding a values ranging from 6 to 8 ns across diferent systems. A value of 10 ns was used to define statistically independent blocks for a block averaging analysis, which confirmed that the block-averaged RMSD values remained stable.

To verify that the simulation results were independent of the initial configuration, RMSD distributions and error estimates were compared across the three replicas of each simulated system. The similarity in RMSD distributions and block-averaged means indicates that all three replicates sampled statistically consistent conformational ensembles. These analyses confirm that the simulations were not biased by their starting structures, and that the observed dynamics are representative of an equilibrated system. Analyses on the conformational ensembles were performed on MD metatrajectories constructed by concatenating the equilibrated part of three independent replicates fo each simulated syste, to enhance sampling of the conformational space. Analyses were carried out using the *cpptraj* program from the *AmberTools* suite (version 24)[8,9].

SUPPLEMENTARY REFERENCES

1. Graphpad Software, L.J., California, USA. Graphpad Prism 7. (Graphpad Software, La Jolla, California, USA). doi:
2. Collaborative Computational Project, N. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**, 760-3 (1994). doi: 10.1107/S0907444994003112
3. Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-7 (2004). doi: 10.1093/nar/gkh340
4. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **42**, W320-4 (2014). doi: 10.1093/nar/gku316
5. Panjkovich, A. & Svergun, D.I. CHROMIXS: automatic and interactive analysis of chromatography-coupled small angle X-ray scattering data. *Bioinformatics* (2017). doi: 10.1093/bioinformatics/btx846
6. Petoukhov, M.V. et al. New developments in the ATSAS program package for small-angle scattering data analysis. *Journal of Applied Crystallography* **45**, 342-350 (2012). doi: 10.1107/S0021889812007662
7. Iacobucci, C. et al. A cross-linking/mass spectrometry workflow based on MS-cleavable cross-linkers and the MeroX software for studying protein structures and protein-protein interactions. *Nat Protoc* **13**, 2864-2889 (2018). doi: 10.1038/s41596-018-0068-8
8. Case, D.A. et al. The Amber biomolecular simulation programs. *J Comput Chem* **26**, 1668-88 (2005). doi: 10.1002/jcc.20290
9. Case, D.A. et al. AmberTools. *J Chem Inf Model* **63**, 6183-6191 (2023). doi: 10.1021/acs.jcim.3c01153
10. Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P.A. & Case, D.A. Development and testing of a general amber force field. *J Comput Chem* **25**, 1157-74 (2004). doi: 10.1002/jcc.20035
11. Jakalian, A., Jack, D.B. & Bayly, C.I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J Comput Chem* **23**, 1623-41 (2002). doi: 10.1002/jcc.10128
12. Tian, C. et al. ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J Chem Theory Comput* **16**, 528-552 (2020). doi: 10.1021/acs.jctc.9b00591
13. M. Bradbrook, G. et al. X-Ray and molecular dynamics studies of concanavalin-A glucoside and mannoside complexes Relating structure to thermodynamics of binding. *Journal of the Chemical Society, Faraday Transactions* **94**, 1603-1611 (1998). doi: 10.1039/A800429C
14. Anandakrishnan, R., Aguilar, B. & Onufriev, A.V. H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res* **40**, W537-41 (2012). doi: 10.1093/nar/gks375
15. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. & Klein, M.L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **79**, 926-935 (1983). doi: 10.1063/1.445869
16. Joung, I.S. & Cheatham, T.E. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *The Journal of Physical Chemistry B* **112**, 9020-9041 (2008). doi: 10.1021/jp8001614
17. Berendsen, H.J.C., Postma, J.P.M., Gunsteren, W.F.v., DiNola, A. & Haak, J.R. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **81**, 3684-3690 (1984). doi: 10.1063/1.448118
18. Miyamoto, S. & Kollman, P.A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry* **13**, 952-962 (1992). doi: 10.1002/jcc.540130805
19. Loncharich, R.J., Brooks, B.R. & Pastor, R.W. Langevin dynamics of peptides: the frictional dependence of isomerization rates of N-acetylalanyl-N'-methylamide. *Biopolymers* **32**, 523-35 (1992). doi: 10.1002/bip.360320508
20. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **98**, 10089-10092 (1993). doi: 10.1063/1.464397