

Materials and Methods and SI Appendix Information

Inter-domain conformational flexibility underpins the activity of UGGT, the eukaryotic glycoprotein secretion checkpoint.

Pietro Roversi^{1‡*}, Lucia Marti^{2‡}, Alessandro T. Caputo^{1‡}, Dominic S. Alonzi^{1‡}, Johan C. Hill^{1‡}, Kyle C. Dent³, Abhinav Kumar¹, Mikail D. Levasseur¹, Andrea Lia^{1,2}, Thomas Waksman¹, Souradeep Basu¹, Yentli Soto Albrecht¹, Kristin Qian¹, James Patrick McIvor¹, Colette B. Lipp¹, Dritan Siliqi⁴, Snežana Vasiljević¹, Shabaz Mohammed¹, Petra Lukacik³, Martin A. Walsh³, Angelo Santino² and Nicole Zitzmann^{1*}

¹: Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, England, United Kingdom. ²: Institute of Sciences of Food Production, C.N.R. Unit of Lecce, via Monteroni, I-73100 Lecce, Italy. ³: Diamond Light Source & The Research Complex at Harwell, Didcot OX11 0DE, England, UK. ⁴: Istituto di Cristallografia (IC-CNR), via Amendola 122/O, I-70126 Bari, Italy.

‡: these authors contributed equally; *: corresponding authors

Materials and methods

*Ct*UGGT cloning, expression and purification.

All DNA primers were synthesized by Sigma Aldrich. All DNA agarose gels were run in TAE containing 1X GelRed (Biotium) electrophoresed in TAE for 45 min at 110V and DNA bands gel-purified using QIAquick Gel Extraction Kit (QIAGEN). DNA minipreps used the QIAprep Spin Miniprep Kit (QIAGEN). DNA maxipreps used either the QIAfilter Plasmid Maxi kit (QIAGEN) or the GenElute HP Plasmid Maxiprep Kit (Sigma-Aldrich).

Amplification of CtUGGT gene DNA from cDNA. Amplification of the mature sequence (residues 24-1505, with the exclusion of the C-terminal KDEL sequence, residues 1506-1509) of the *Chaetomium thermophilum* UGGT gene *aka CtUGGT* (CTHT-0048990, UniProt entry G0SB58_CHATD) from a *Ct* cDNA library (1) was accomplished in two rounds. First, 25 µL of 2X Q5 high fidelity master mix (New

England Biolabs (NEB)) were added to 2.5 μ L of 10 μ M forward and reverse primers 1F and 1R (SI Appendix Table S5), 1 μ L of 11 ng/ μ L *Ct* cDNA and 19 μ L of water; PCR protocol: step 1: 98 $^{\circ}$ C for 30 s.; step 2: 98 $^{\circ}$ C for 10 s; step 3: a gradient from 62 to 66 $^{\circ}$ C in 20 s; step 4: 72 $^{\circ}$ C for 1 min 30s; steps 2-5 were repeated 5 times; step 6: 98 $^{\circ}$ C for 10 s; step 7: 70 $^{\circ}$ C for 20 s; step 8: 72 $^{\circ}$ C for 1 min 30s; steps 6-8 were repeated 25 times; step 10: 72 $^{\circ}$ C for 1 min 30s. This first round of amplification resulted in 30 μ L of 19.7 ng/ μ L of *CtUGGT* gene DNA, 1 μ L of which was used as a template for a second round of amplification (PCR protocol as listed above for the first amplification except for the template). The PCR amplification mix was loaded on a 1% agarose gel and the bands corresponding to the ~4500 bp *CtUGGT* gene DNA were excised and purified and concentrated to 45.1 ng/ μ L by ethanol precipitation.

Insertion of the CtUGGT gene into the pPICLIT expression vector. *E. coli* NEB 5 α competent cells were transformed with the pPICLIT vector (derived from pPICZ α -A (ThermoFisher Scientific by digesting a LITMUS28i vector (NEB), isolating its multiple cloning site and ligating it into pPICZ α -A 's multiple cloning site; pPICLIT harbors a C-terminal hexa-histidine tag and carries a zeocin resistance cassette). The cells transformed with pPICLIT were grown overnight at 37 $^{\circ}$ C in low-salt LB broth 25 μ g/mL zeocin to an OD₆₀₀ of about 3.7. The cells were pelleted by centrifugation at 4000 *g* and the pellet used for preparing 500 μ L of plasmid at 18.7 ng/ μ L. This sample was digested with 1 μ L of *Xho*I and 1 μ L *Xba*I restriction enzymes (NEB) at 20000 units/mL at 37 $^{\circ}$ C for two hours. The digested vector was then run on a 1% agarose gel and gel-extracted giving 10.2 ng/ μ L digested pPICLIT plasmid. This was concentrated to 44.1 ng/ μ L of DNA using ethanol precipitation. To ligate the *CtUGGT* gene DNA to the pPICLIT vector, Gibson assembly (NEB) was used. Briefly, 10 μ L of 2X Gibson assembly master mix (NEB) were mixed with 3 μ L of 44.1 ng/ μ L pPICLIT vector, linearized as described above, 7.5 μ L of 45.1 ng/ μ L *CtUGGT* gene DNA amplified as described above. The mixture was incubated in a thermocycler at 50 $^{\circ}$ C for 15 min to assemble. 2 μ L of this Gibson assembly mixture was used to transform *E. coli* NEB 5 α competent cells; after overnight growth the *CtUGGT*:pPICLIT plasmid was purified to a final concentration of 15 ng/ μ L and DNA-sequenced with AOX1 forward and reverse sequencing primers 2F and 2R (Source Bioscience), see SI Appendix Table S5.

Deletion of a DNA 64-bp intron in the cDNA-amplified CtUGGT gene DNA. Primers 3F and 3R, complementary to either side of the insert, were designed using the NEB

Base Changer server at NEB (see SI Appendix Table S5). The mutagenesis reaction was set up by mixing 12.5 μ L of 2X Q5 high fidelity master mix (NEB), 1.25 μ L of 10 μ M of forward and reverse primers, 1 μ L of 15 ng/ μ L *CtUGGT*:pPICLIT plasmid and 9 μ L of water. The following PCR protocol was effected: step 1: 98 °C for 30 s; step 2: 98 °C for 10 s; step 3: 67 °C for 20 s; step 4: 72 °C for 4 min; repeat steps 2-4 25 times; step 6: 72 °C for 30 s. The methylated template DNA was then digested and the PCR-amplified mutated *CtUGGT*:pPICLIT DNA circularized by adding 1 μ L of the PCR product to 5 μ L of 2X KLD reaction buffer (NEB), 1 μ L of 2X KLD enzyme mixture (NEB) and 3 μ L of water. This mixture was incubated for 5 min at room temperature and then used to transform TOP-10 competent cells (ThermoFisher Scientific); the mutated *CtUGGT*:pPICLIT plasmid from these cells was minipreped and DNA-sequenced with standard AOX1 forward and reverse sequencing primers 2F and 2R, see SI Appendix Table S5 (Source Bioscience).

Insertion of the CtUGGT gene into the pHLsec expression vector. Cloning of *CtUGGT* into the vector pHLsec for mammalian cell secreted expression (2) was again accomplished using the Gibson assembly protocol (NEB). The plasmid was linearized by digestion with High-Fidelity *KpnI* and *AgeI* restriction enzymes (NEB) using CutSmart Buffer (NEB) in 20 μ L reaction volumes for 1 hour at 37 °C according to the manufacturer's instructions. PCR amplification of the *CtUGGT* gene for insertion into linearized pHLsec used the *CtUGGT*:pPICLIT plasmid as template with forward and reverse primers 4F and 4R (SI Appendix Table S5) and Q5 Hot Start High-Fidelity DNA Polymerase (NEB). DNA bands corresponding to linearization and amplification products were gel-purified from 0.8% agarose gels with the GenElute PCR Clean-up kit (Sigma-Aldrich). The Gibson assembly reaction was carried out incubating for 1 hour at 50 °C a 20 μ L total volume reaction containing 76 ng of linearized pHLsec vector and 144 ng of *CtUGGT* DNA PCR-amplified insert. NEB 5 α competent *E. coli* cells were transformed and plated onto LB agar + Carbenicillin (5 μ g/mL) plates for ~ 16 hours. Plasmid DNA was minipreped and DNA-sequenced at the vector-insert junctions with forward and reverse primers 5F and 5R (SI Appendix Table S5) and DNA for transfection was maxipreped from frozen aliquots of *E. coli* colonies harboring *CtUGGT*:pHLsec and containing 20% glycerol.

*Cloning of the double mutants *CtUGGT*^{N796C/G1118C} and *CtUGGT*^{D611C/G1050C}.* The coding sequence of *CtUGGT* (residues 24-1505, previously cloned in pHLsec) was

amplified by PCR using the Accuprime Pfx DNA Polymerase (ThermoFisher Scientific) and cloned into the pDONR221 vector (using primers 6F, 6R, see SI Appendix Table S5). The QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent) was used to introduce the mutations G1118C, N796C (primers 7F, 7R and 8F, 8R in SI Appendix Table S5), and G1050C, D611C (primers 9F, 9R and 10F, 10R in SI Appendix Table S5), the protocol followed the manufacturer's instructions. The DNA double mutants *CtUGGT*^{N796C/G1118C} and *CtUGGT*^{D611C/G1050C} in pDonr221 were amplified by PCR with primers 4F and 4R (SI Appendix Table S5) and inserted in pHLsec by Gibson assembly as described for the native sequence above.

CtUGGT protein expression in HEK293F cells. Human epithelial kidney FreeStyle 293F cells (ThermoFisher Scientific) at 10⁶ cells/mL suspended in FreeStyle 293 Media (ThermoFisher Scientific) were transfected using the FreeStyle 293 expression system (ThermoFisher Scientific). For a 50mL culture: 62.5 µL of FreeStyle MAX transfection reagent (ThermoFisher Scientific) and 62.5 µg of plasmid DNA were each separately diluted to 1 mL with OptiPRO SFM reagent (ThermoFisher Scientific) then mixed, incubated for 10 min at room temperature and finally added to the cell suspension. Transfected cells were left shaking in 500 mL Erlenmeyer flasks with 0.2 µm vent caps (Corning) shaking at 135 revolutions per min (rpm) in a 37 °C incubator kept at 8% CO₂. Protein expression was checked by taking 1 mL aliquots of cells every 24 hours for a week and analyzing the time course on an 4-12%BisTris gel (ThermoFisher Scientific). The double mutants *CtUGGT*^{N796C/G1118C} (“intermediate”) and *CtUGGT*^{D611C/G1050C} (“closed”) were expressed in similar manner to the native *CtUGGT*.

CtUGGT protein purification. The HEK293F cells were spun by centrifugation for 30 min at 1,000 g and the supernatant made to 1X Phosphate Buffer Saline (PBS) by addition of a 10X PBS stock. The solution was syringe-filtered through a 0.2 µm filter and applied at a flow rate of 1 mL/min (at 4 °C or at 21 °C) onto a 1 mL HisTrap HP Ni-affinity column (GE Healthcare) equilibrated against PBS binding buffer. The column was washed with 20 mL of binding buffer and the protein eluted with a linear gradient of 40 mL going from binding buffer to the same buffer made 400 mM imidazole at a rate of 1 mL/min. The fractions containing *CtUGGT* were pooled and loaded onto a HiLoad Superdex 200 16/60 column size-exclusion chromatography column (GE Healthcare) equilibrated against the running buffer: 20 mM HEPES pH 7.4, 140 mM NaCl. The size-exclusion chromatography experiment was run at 0.5

mL/min and at 21 °C; the elution fractions containing the protein were concentrated by PES-membrane 50 kDa MW-cutoff spin-concentrators (Sartorius). The double mutants *CtUGGT*^{N796C/G1118C} and *CtUGGT*^{D611C/G1050C} were purified in similar manner to the native *CtUGGT*. The SEC-MALS experiment injected 100 µL of 0.94 mg/mL protein onto a Superdex 200 10/300 GL column size-exclusion chromatography column (GE Healthcare) running at 0.5 mL/min in PBS running buffer. Samples were analyzed using a Prominence HPLC (Shimadzu) at 0.5 ml/min with an online UV, refractive index and a Dawn HELEOS 8+ (Wyatt Technologies) multi-angle laser light scattering detector set to 662.3 nm. Peaks were analyzed with a Zimm model using a refractive index increment of 0.185 ml/g.

CD spectroscopy.

The Circular Dichroism data were collected on a Jasco CD spectrophotometer in the wavelength range of 185 nm to 260 nm. The data between 185 and 240 nm were analyzed with the Dichroweb server using the CDSSTR method with reference set 3 (3). 50 µL of protein at 1.0 mg/mL was dialyzed three times against 1 l of fresh 20 mM potassium phosphate pH 7.4 for 2 hours at 4 °C using a 10 kDa MWCO Slide-A-Lyzer (ThermoFisher Scientific) dialysis capsule, and then diluted in the same buffer to a final volume of 200 µL and a concentration of 0.1 mg/mL. CD measurements were taken on a J-815 CD instrument (Jasco) with an accessory CDF-426S thermoregulator (Jasco). Measurements were taken in a 1 mm QS polarized quartz cuvette (Hellma Analytics) and were averaged over three accumulations. Melt curves were collected at 2 °C intervals between 20 °C and 90 °C with Data Pitch = 0.5 nm, Data Integration Time = 1.00 s, Bandwidth = 1.00 nm, and scanning speed = 50 nm/min. The curves obtained were cut if high tension (HT) voltage exceeded 600 V after which the instrument was considered to be saturated. CD spectra were smoothed using the Savitzky-Golay method with convolution width of 15.

***CtUGGT* activity assay: glucosylation of urea-denatured thyroglobulin.**

Bovine thyroglobulin (Sigma-Aldrich) was denatured with urea following the protocol by Trombetta *et al* (4). Each reaction mixture contained 100 µg of urea-denatured bovine thyroglobulin, 86 µM UDP-Glucose, 8.6 mM CaCl₂, 8.6 mM Tris-HCl pH 8.0 and 45 pmol of *CtUGGT* enzyme (either wild-type, or one of the double Cys mutants). The reaction mixtures were set up at 37 °C. Each reaction was 70 µL to

start with, in triplicate. 10 μ L aliquots were taken at each time point (5', 15', 30', 1 h, 2 h and O/N), and the glucosylation quenched by addition to each 10 μ L aliquot of 1 μ L of PNGaseF denaturing buffer, then heating for 10 min at 90 °C. Then 5 μ L of 10X PNGase glycobuffer 2 (NEB), 5 μ L of NP40 10%, 1 μ L of PNGase F at 1 mg/mL and 27 μ L of water were added to each sample for the overnight digestion with PNGase F. The *N*-linked glycan were labelled with anthranilic acid (2-AA) (Sigma-Aldrich), purified by adsorption to Speed-amide SPE columns and detected by normal-phase high-performance liquid chromatography, as described by Caputo *et al* (5).

The amount of glucosylation was measured in comparison to control by measuring the peak area of the PNGase F released 2-AA-labelled species Man₉GlcNAc₂ and Glc₁Man₉GlcNAc₂ using Waters Empower software. This allows the % of glucosylation to be determined as the % of Glc₁-species (Peak Area Glc₁Man₉GlcNAc₂) as a total of potential glucosylation species (Peak Area of Glc₁Man₉GlcNAc₂ + Man₉GlcNAc₂).

Small Angle X-ray Scattering (SAXS)

SAXS was measured on beamLine BM29 at the ERSF. Briefly, CtUGGT at 2.0 mg/mL in 10 mM HEPES pH 7.4 150 mM NaCl and 5% glycerol was cleared by centrifugation at 16000 *g* for 5 min and then spun for 5 min at 12000 *g* through a PVDF UltraFree 0.1 μ m spin filter (Millipore). The buffer (obtained from the size-exclusion chromatography after one column volume was eluted) was spin filtered through the same filter and used to set up a dilution series of the protein, from 2.4 to 0.06 mg/mL. For each sample, ten scans of 1 s each were acquired measuring solution scattering at 20 °C with an X-ray beam of $\lambda=0.9919$ Å using a flow rate of 3 μ L/s through a 1.8 mm diameter quartz capillary sample cell. Buffer scattering was measured before and after each series of 10 scans in the same conditions as the protein samples. The scattering was recorded by a Pilatus 1M detector. Data were processed with the ScÅtter and the ATSAS suite (6).

Electron Microscopy

Cryo-microscopy specimens were prepared by applying 3 μL of CtUGGT (0.25 mg/mL in 20 mM HEPES, 140 mM NaCl, pH 7.3) to previously glow-discharged UltrAufoil 2/2 gold specimen grids (Quantifoil R2/2, Quantifoil Micro Tools GmbH). Specimens grids were blotted for 2.5 seconds, and immediately plunged into liquid ethane using a Vitrobot Mark IV (FEI Company). Electron micrographs were recorded on a Titan Krios microscope (FEI Company) at the electron bio-imaging centre (eBIC), Diamond Light Source (Harwell, Oxfordshire, UK). The microscope was operated at 300 kV using a Volta phase plate. Exposure movies consisting of 20 frames were recorded using a Gatan K2 Summit detector (Gatan, Pleasanton, CA). Micrographs were recorded using a total exposure of 10 seconds (total dose of 55 $\text{e}^-/\text{\AA}^2$) at an object pixel size of 1 $\text{\AA}/\text{pixel}$. A nominal defocus of 700 nm was applied over the duration of the collection. Movie frames were aligned using the ‘optical flow’ function of the SCIPION package (7). Micrograph CTF parameters were determined using CTFFIND 4.1 (8). Micrographs demonstrating poor particle distribution, ice thickness or unacceptable motion were excluded. A subset of particles was initially selected manually and subjected to 2D classification in RELION 1.4. The remaining particles in 150 micrographs were selected by template matching using RELION. 88,567 particles were down sampled to 4 $\text{\AA}/\text{pixel}$ and subjected to 3 rounds of 2D classification (RELION) in which bad classes were manually excluded from the dataset. 33,819 images, contributing to classes with the characteristic appearance of CtUGGT were used to generate an *ab initio* three-dimensional (3D) starting model using the PRIME algorithm (9) of the SIMPLE package (10); this process converged reliably after 10 iterations. One of the two hands of the *ab initio* model showed strong similarity to the CtUGGT crystal structure, supported by a convincing fit for all but for the TRXL2 domain, and was used as a reference for 3D classification into 4 classes in RELION. This 21 \AA reconstruction and orientation distribution strongly suggested that multiple orientations of TRXL2 were interfering with particle alignment. Attempts at 3D classification to reveal discrete conformations for TRXL2 failed. TRXL2, forming a sizeable contribution to the shape of the particle, was flexible enough to prevent successful alignment of the particles in the absence of a tight mask around the core complex. Exclusion of TRXL2 from the mask used for alignment was necessary in order to see the resolution

improvement from 21 to 15 Å. Docking of the crystal structure in the cryo-EM envelope was done in Chimera (11).

Mass spectroscopy: tryptic peptides

Tryptic peptides of CtUGGT, and the double mutants CtUGGT^{N796C/G1118C} (“intermediate”) and CtUGGT^{D611C/G1050C} (“closed”) were separately run on an EASY-nLC43 1000 UHPLC system (Proxeon) and electrosprayed directly into a Q Exactive mass spectrometer (Thermo Fischer Scientific) through an EASY-Spray nano-electrospray ion source (Thermo Fischer Scientific). Peptides were trapped on an in-house packed guard column (75 µm i.d. x 20 mM, repositil C18, 3µm, 120 Å) using solvent A (0.1% Formic Acid in water) at a pressure of 500 bar and then separated using an EASY-spray Acclaim PepMap® analytical column (75 µm i.d. × 500 mm, RSLC C18, 2 µm, 100 Å) eluted with a linear gradient (7 % to 31 % solvent B (0.1% formic acid in acetonitrile) in 60 min) at a flow rate of 200 nL/min. Full scan MS spectra were acquired in the Orbitrap (scan range 350-2000 *m/z*, resolution 70000, AGC target 3e6, maximum injection time 100 ms). After the MS scans, the 10 most intense peaks were selected for HCD fragmentation at 30 % of normalized collision energy. HCD spectra were also acquired in the Orbitrap (resolution 17500, AGC target 5e4, maximum injection time 120 ms) with first fixed mass at 120 *m/z*.

Data analysis (crosslinking): MS data were converted into .mgf format using MSconvert from the ProteoWizard toolbox (12) and searched using the pLink software (13). The database contained CtUGGT only. Search parameters were as follows: maximum number of missed cleavages = 2, fixed modification = none, variable modification 1=Oxidation-Met, variable modification 2=carbamidomethyl-Cys, disulphide bonds were considered as the crosslink (Cys-Cys, -2.015650 Da), mass accuracy filter = 20 ppm for precursor ions with consideration of the first 5 isotopic peaks, MS2 tolerance = 20 ppm. Data were filtered by E-value ($E < 1.0 \times 10^{-7}$) and a manual inspection was performed.

Data analysis (protein identification). The raw data files generated were processed using MaxQuant software (Version 1.5.0.35), integrated with Andromeda search engine as previously described (14). Raw MS data were searched against database containing proteins of interest as well as list of common contaminants by Andromeda. Trypsin with a maximum number of 2 missed cleavages was selected as the protease.

Acetylation (Protein N-term), Carbamidomethylation (Cysteine) and Oxidation (Methionine) were used as variable modifications. Protein and PSM false discovery rate (FDR) were set at 0.01.

Mass spectroscopy: peptic peptides.

CtUGGT^{N796C/G1118C} (“intermediate”) and *CtUGGT*^{D611C/G1050C} (“closed”) samples were digested in-solution with sequencing grade pepsin (Promega) in accordance to the manufacturers’ instructions. Briefly, samples were treated with 100 mM IAA for 1 hour (in the dark) to alkylate any free cysteines followed by denaturing the samples using 8 M urea in 1% TFA in water (v/v) for 40 minutes. Next 360 µL of 1% TFA in water (v/v) was added to reduce the urea concentration to 1 M. Pepsin was reconstituted to 300 ng/ml with 5mM ammonium bicarbonate and 300 ng (1 µL) was added to 2 µL (3 µg/µL) *CtUGGT*^{N796C/G1118C} and *CtUGGT*^{D611C/G1050C} with 2 µL of 5% trifluoroacetic acid (TFA) and incubated overnight at 37 °C. Samples were analyzed by reversed phase LC-MS using a Dionex UltiMate3000 RSLC (Thermo Scientific) coupled to a Q-Exactive (Thermo Scientific). 5 µL (75 ng on column) of digested sample was desalted using a C18 PepMap trapping column (µ-Precolumn, 300 µM I.D. x 5 mm, 100 µm particle size, 100 Å, Thermo Scientific) at a flow rate of 10 µL/min. The trapping buffer was 0.05% v/v trifluoroacetic acid (TFA) in water (LC-MS grade). Samples were then separated using a C18, 75 µm x 15 cm (2.6 µm particle size, 150 Å; Thermo Scientific) analytical column at 300 nL/min with a 40 minute linear gradient as follows: 0-4 (min) 3% B, 4-10 (min) 8% B, 10-40 (min) 35% B, 40-41 (min) 85% B, 41-46 (min) 85% B, 46-47 (min) 3% B, 47-60 (min) 3% B. Mobile phase A was 0.1% v/v formic acid in water (LC-MS grade) and B was 0.1% v/v formic acid in acetonitrile (LC-MS grade). Data was acquired in the Data Dependent Mode (DDA) using the following settings: chromatographic peak width: 12 s, resolution: 70,000, AGC target: 3e6, maximum IT (injection time): 100 ms, scan range: 300 to 2000 m/z. The dd-MS2 conditions - resolution: 17,500. The AGC target conditions- 5e4, Maximum IT: 100 ms, loop count: 10 (i.e. Top 10), isolation width: 2.0 m/z, fixed first mass: 120.0 m/z, and the data dependent (dd) settings -under fill ratio: Minimum AGC target: 3.0 e3, intensity threshold: 5.0 e4, charge exclusion: 1 and >8, peptide match: preferred, exclude isotope: on, dynamic exclusion: 30.0 s. A normalized collision energy (NCE) of 27 was used for the fragmentation of peptides in a high-energy collision dissociation (HCD) cell and the s-lens setting in the tune file

was changed to 70. A normalized collision energy (NCE) of 27 was used for the fragmentation of peptides in a high-energy collision dissociation (HCD) cell and the s-lens setting in the tune file was changed to 70. Raw data was converted to MGF files with MSconvert (64-bit) (15) and the presence of disulfide bonds was analyzed with MassMatrix software (16).

Protein crystallization and crystals X-ray diffraction

All sitting drops were set up either with either 100 nL or 200 nL of protein and 100 nL of mother liquor solution, and equilibrated against 50 μ L of mother liquor solution at 21 °C. All crystals were harvested and flash-frozen in liquid N₂ (the mother liquor solutions were cryo-protected already).

P6₁ crystal form: 2 μ g of a glutathione-S-transferase (GST) fusion of Endo F1 glycosidase were added to 200 μ L of CtUGGT in 20 mM HEPES pH 7.4, 150 mM NaCl at OD₂₈₀=6.95 *i.e.* 6.0 mg/mL (calculated absorption coefficient ϵ_{280} =1.13 (mL/mg)⁻¹ cm⁻¹). The mother liquor was designed to screen around initial crystal hits in condition E2 of the Morpheus Screen (17) (Molecular Dimensions): 51.4 mM MES pH 3.3, 44.6 mM Imidazole pH 8.3 (pH~7.0), Morpheus precipitant mix 2 50.2% v/v, 120 mM Morpheus Ethylene Glycols mix. For experimental phasing, the crystals were soaked in 10mM K₂PtBr₆, or 10mM K₂PtCl₆ or 10mM K₂PtI₆ for 10 min. One crystal was set to soak in 10mM K₂PtI₆ immediately before a fire alarm went off and we had to leave the building. As a result, that K₂PtI₆ soak lasted 45 minutes. That crystal was the one that enabled structure solution.

P6₁22 crystal form: CtUGGT in 10 mM HEPES pH 8.0, 150 mM NaCl, at OD₂₈₀=16.0 *i.e.* 14.0 mg/mL, was mixed 2:1 with and set to equilibrate against 20% of condition D12 of the Morpheus screen (17) (Molecular Dimensions), *i.e.* 24 mM Morpheus Alcohols mix, 20 mM Morpheus Buffer System 3 pH 8.5, 2.5% v/v MPD; 2.5% w/v PEG 1000, 2.5% w/v PEG 3350, supplemented with 1.6 mM CaCl₂ and 50 μ M UDP-glucose.

P4₃ crystal form: CtUGGT in 20 mM sodium citrate pH 6.0, 140 mM NaCl, 0.1 mM CaCl₂ and MnCl₂, 0.15 mM UDP-Glc, 0.1 mM Octaethylene glycol monododecyl ether (C12E8, from Sigma), and 5% v/v glycerol, at OD₂₈₀=6.0 *i.e.* 5.25 mg/mL, was mixed 1:1 with and set to equilibrate against condition H8 of the Morpheus Screen (Molecular Dimensions): 0.1 M Morpheus Amino acids mix, 0.1 M Morpheus Buffer

System 2 pH 7.5, 50 % v/v Morpheus Precipitant Mix 4, *i.e.* 12.5% v/v MPD; 12.5% w/v PEG 1000, 12.5% w/v PEG 3350. $P2_12_12_1$ CtUGGT^{D611C/G1050C}: CtUGGT^{D611C/G1050C} at OD₂₈₀=4.3 *i.e.* 3.8 mg/mL in 50 mM MES pH 6.5, 50 mM NaCl, 1.0 mM CaCl₂, 0.25 mM UDP was mixed 2:1 with condition G2 of the Morpheus Screen (Molecular Dimensions) (17) and set to equilibrate against it: 0.1 M Morpheus Carboxylic acids mix, 0.1 M Morpheus buffer system 1 pH 6.5, 50 % v/v Morpheus precipitant mix 2, *i.e.* 20.0% v/v ethylene glycol; 10.0% w/v PEG 8000.

X-ray diffraction data from the $P6_1$ crystals were collected on BM14 at the ERSF, Grenoble, France. The native data were collected at a wavelength $\lambda=0.97880$ Å. A fluorescence spectrum across the Pt L-III edge was then collected from the $P6_1$ K₂PtI₆ soaked crystal, and analyzed by Chooch (18) to give $\lambda_{\text{peak}}=1.07141$ Å, at which wavelength diffraction data from all the Pt-salt soaked crystals were collected. At this wavelength, anomalous and dispersive contributions to Pt scattering were $f'_{\text{Pt}}=24.3$ e⁻ and $f''_{\text{Pt}}=-17.8$ e⁻. X-ray diffraction data from the $P4_3$ and $P6_122$ crystal forms were collected on beamline I04-1 at the Diamond Light Source Harwell England UK, at wavelengths $\lambda=0.92821$ Å and $\lambda=0.91741$ Å, respectively. X-ray diffraction data from the $P2_12_12_1$ CtUGGT^{D611C/G1050C} crystal were collected on beamline ID30A-1 at the ESRF, Grenoble, France, at wavelength $\lambda=0.96600$ Å.

All X-ray diffraction images were indexed with XDS (19) run from the automated processing suite autoPROC and scaled with CCP4 Aimless (20). Anisotropy correction was carried out using the Global Phasing STARANISO server at <http://staraniso.globalphasing.org/cgi-bin/staraniso.cgi>.

Phase determination, model building and structure refinement.

$P6_1$ crystal form. Six Pt sites were found by SHELXD (21), run from the autoSHARP suite of programs (22), analyzing in $P6_1/P6_5$ the anomalous differences of the 3.5 Å diffraction data from the K₂PtI₆ soaked crystal. Completion and refinement of the heavy atom model was carried out in SHARP (23) against the MIR-AS data collected from the native and the three Pt salt soaks (data collection statistics in Table 1, phasing statistics in Table 2). Only the Pt sites occupancies were refined, while the Pt f' and f'' values were kept fixed at the experimentally measured values and their B factors were kept fixed at the B_{Wilson} of the crystals. Initial phases were good to about

6.2 Å. Solvent flattening to 3.5 Å in Solomon confirmed $P6_1$ as the correct spacegroup ($CC_{P6_1}=57.8\%$ vs. $CC_{P6_5}=45.7\%$) and yielded an electron density map in which most of the main chain was clearly traceable. Phased molecular replacement in CCP4-Molrep (24) using the solvent flattened phases allowed correct positioning of PDB ID 3WZS (the TRXL3 domain of *CtUGGT* (25)), of a TRXL2 homology model based on PDB ID 3BCI (*Staphylococcus aureus* DsbA) at the TRXL2 location, and of a copy of PDB ID 1G9R (the bacterial galactosyltransferase LgtC) at the GT domain location. Additional strands and helices visible in the map were fitted to the solvent flattened map in Coot (26) and an initial main-chain model iteratively built alternating refinement in autoBUSTER (27) and CNS-DEN (28), with one TLS group per domain and secondary structure restraints throughout and manual model building in Coot. Automated model building and solvent flattening in Phenix MR-Rosetta (29) produced density modified electron density maps which showed density for many side chains and were initially used to dock the sequence in place. The final refinements were carried out alternating Phenix (30) refinements with TLS, rotamer and secondary structure restraints and autoBUSTER (27) refinements with TLS and external restraints (31) to the $P2_12_12_1$ *CtUGGT*^{D611C/G1050C} 2.8 Å crystal structure (see below). Ramachandran plot: 93.9% (1288/1371) of all residues were in favored regions; 99.1% (1359/1371) of all residues were in allowed regions; 0.9% (12/1371) residues were Ramachandran outliers.

WT P6₁22, P4₃ and CtUGGT^{D611C/G1050C} *double mutant P2₁2₁2₁ crystal forms.* These crystal forms were initially phased by molecular replacement in Phaser (32), searching with the $P6_1$ model without the TRXL2 domain. In each crystal form, the model for TRXL2 was then fitted into the Fo-Fc difference electron density map calculated after rigid body refinement of the 6 individual domains from the initial MR solution. The *CtUGGT*^{D611C/G1050C} $P2_12_12_1$ model was refined in autoBUSTER with TLS restraints. The $P6_122$ and $P4_3$ full models were refined by alternating Phenix refinement with TLS, rotamer and secondary structure restraints and autoBUSTER refinement with TLS and external restraints (31) to the *CtUGGT*^{D611C/G1050C} $P2_12_12_1$ model throughout. In the $P4_3$ crystal form NCS restraints (31) were also used. Ramachandran plot *CtUGGT*^{D611C/G1050C} $P2_12_12_1$: 95.6% (1311/1371) of all residues were in favored regions; 99.7% (1367/1371) of all residues were in allowed (>99.8%) regions; 0.3 % (4/1371) of all residues were Ramachandran outliers. Ramachandran

plot P6₁₂₂: 98.0% (1293/1363) of all residues were in favored regions; 99.6% (1357/1363) of all residues were in allowed (>99.8%) regions; 0.4 % (6/1363) of all residues were Ramachandran outliers. Ramachandran plot P4₃: 86.0% (2349/2731) of all residues were in favored regions; 98.5% (2689/2731) of all residues were in allowed regions; 1.5% (42/2731) of all residues were Ramachandran outliers.

Generation of transgenic plants

CtUGGT was fused at the N-terminal with the monomeric RFP (mRFP) under the control of the 35S promoter in the pH7WGR binary vector, using the Gateway Recombination Cloning Technology (ThermoFisher Scientific). The pH7WGR binary confers resistance to hygromycin to the plant.

Binary vectors containing 35S::RFP-*CtUGGT* and 35S::RFP were amplified in *Escherichia coli* and used for *Agrobacterium tumefaciens* (strain GV3101) transformation. *Arabidopsis thaliana ebs1-3bri1-9* mutant plants were transformed by the floral dip method (33) and transformants were selected on Murashige and Skoog (MS) basal medium (34) containing hygromycin (Hyg; final concentration 25 µg/mL), 1% (w/v) sucrose, 0.7% (w/v) plant agar, pH 5.7. At least 15 independent transgenic T1 plants were selected among those which showed fluorescence in seedling cotyledon cells, as determined by laser-scanning confocal microscopy (LSCM). Three independent T1 lines were selected for the control construct 35S::RFP. Transient co-expression of 35S::RFP-*CtUGGT* and the endoplasmic reticulum marker (ER-YK (35)) was performed in 4-5 weeks old tobacco leaf epidermal cells (*Nicotiana tabacum* plants, cv Petit Havana) according to established protocols (36, 37). *Plant growth conditions*. Seedlings and plants of *Arabidopsis thaliana* were grown at 22 °C and 70% relative humidity under a 16 h light/8 h dark cycle (light intensity approximately 120 µmol/m²s). Tobacco plants were grown at 25 °C and 70% relative humidity under a 14 h light/10 h dark.

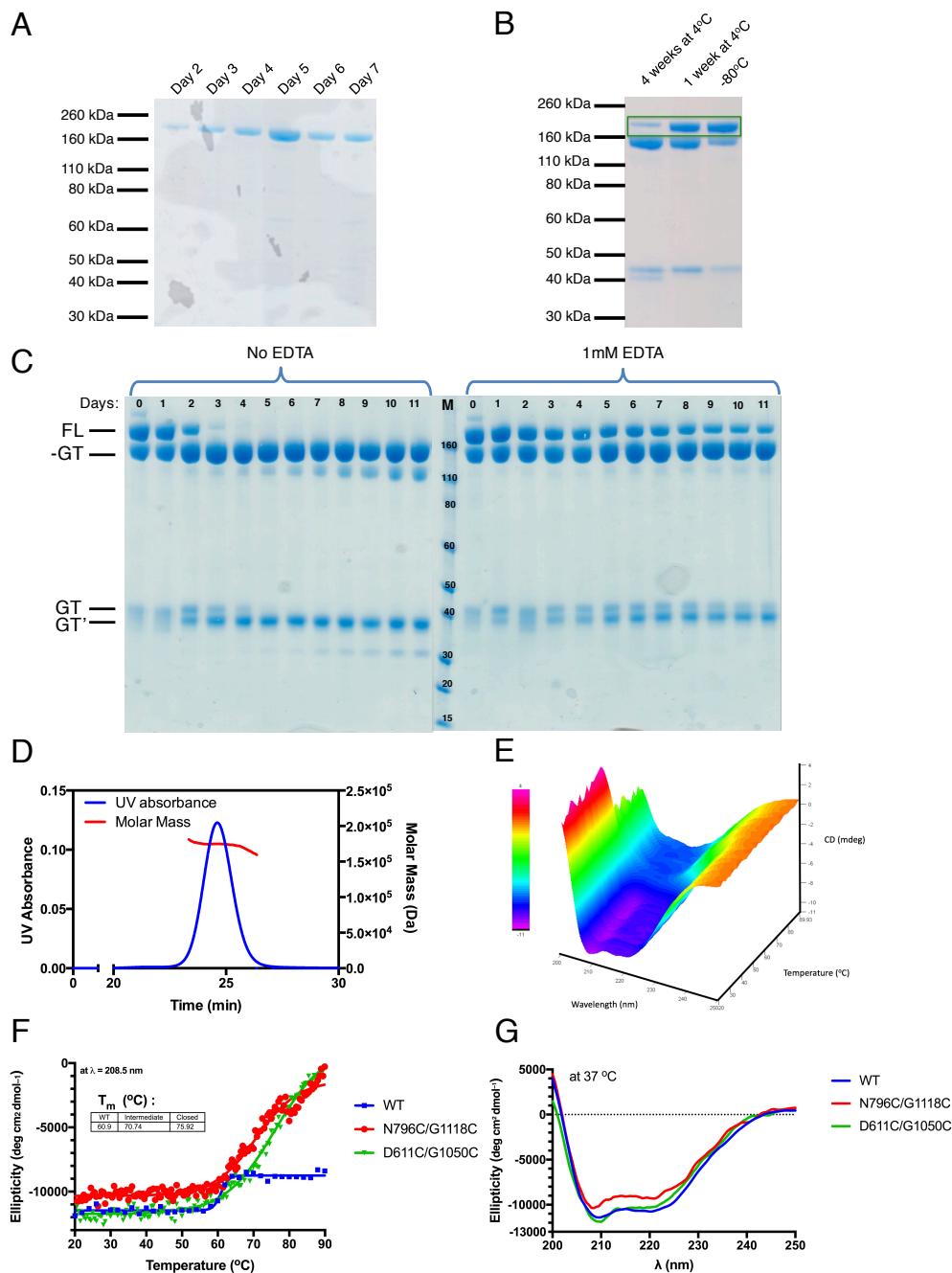
Confocal Laser Scanning Microscopy. For confocal microscopy, seedlings were grown for 10 days in Petri dishes containing one-half-strength MS basal salts medium supplemented with 1% (w/v) sucrose, containing Hyg (25 µg/mL), 0.7% (w/v) plant agar, pH 5.7. Cotyledons were then detached and mounted on a slide. Confocal microscopy was performed using an inverted laser scanning confocal microscope (LSM Pascal; Carl Zeiss). Imaging of 35S::RFP-*CtUGGT* and 35S::RFP were performed using 543 nm He/Ne laser (5 mW). Imaging of ER-YK (ABRC-

Arabidopsis Biological Resource Centre), was performed using 488 nm excitation of an Argon ion laser (25 mW) for YFP. RFP was detected with a 560–615 nm filter set and YFP with a 505–530 nm filter set. Imaging was performed using the objective 40x Zeiss plan-neofluar oil, 1.3 numeric aperture, differential interference contrast. All images are representative of at least five independent experiments.

Western blotting. Microsomal fractions were prepared from adult leaves (approximately 1 g) of 4-week-old plants, as described (38). Proteins extracted were analyzed by western blotting. RFP was imaged using an anti-RFP monoclonal antibody (Chromotek).

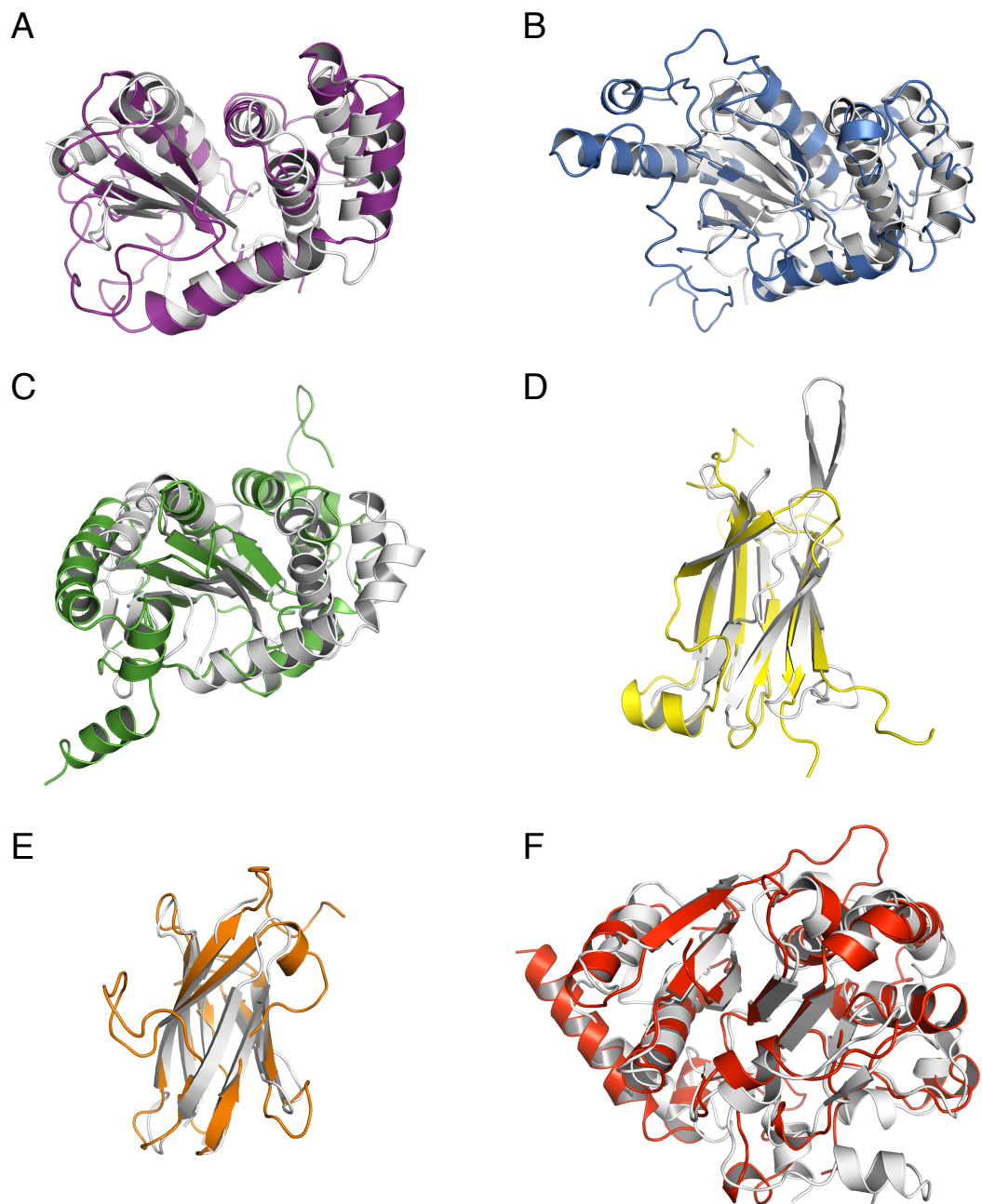
HsUGGT1 and HsUGGT2 Homology Modelling. The *HsUGGT1* and *HsUGGT2* sequences were sourced from the UniProt server and aligned to the *CtUGGT* one using the Clustal Omega server (<http://www.ebi.ac.uk/Tools/msa/clustalo/>). Using this alignment and the *CtUGGT* crystal structure, homology models for *HsUGGT1* and *HsUGGT2* were built using Modeller (39).

Sequence Conservation Mapping. Two multiple sequence alignments were completed using the Clustal Omega server (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) for 20 different species, separately for UGGT1 and UGGT2, sequences sourced from the UniProt server. The sequence conservation patterns across UGGT1 and UGGT2 sequences were mapped onto the structure of *CtUGGT* using the ConSurf server (<http://consurf.tau.ac.uk/2016/> (40)), query sequence *HsUGGT1* and *HsUGGT2* respectively. The conservation mapping was coloured in Pymol, with the scale light blue to dark blue = variable to conserved.



SI Appendix Figure S1. *CtUGGT* expression, purification, endo-proteolysis, circular dichroism and thermal denaturation of wild type *CtUGGT* and *CtUGGT* double Cys mutants. The mature sequence of *CtUGGT* (CTHT_0048990, Uniprot entry G0SB58) comprising residues 24-1505 and omitting the C-terminal ER-retrieval sequence ¹⁵⁰⁶KDEL¹⁵⁰⁹) was cloned into the secreted expression vector

pHLsec (2). Transfection of HEK293F cells with the *CtUGGT*:pHLsec plasmid produced recombinant *CtUGGT*, purified as described in the material and methods. (A) Timecourse of the HEK293F expression of *CtUGGT*, the cloned construct has a theoretical mass of 169,414 Da and the 5 N-linked glycans increase the mass to ~174,00 Da. (B) Recombinant *CtUGGT* quickly undergoes endo-proteolysis: the cleavage occurs in a low sequence identity region between the β S2 and GT domains. N-terminal sequencing reveals two distinct 45 kDa C-terminal fragments starting with the sequences ¹¹⁶⁶NLVSR¹¹⁷⁰ and ¹¹⁷¹GIKFA¹¹⁷⁵ which are located ~30 residues upstream of the beginning of the ordered structure of the GT domain in the crystals. (C) The endo-proteolysis is divalent-metal dependent. FL=Full Length. GT=Glucosyl Transferase domain. –GT: *CtUGGT* without the GT domain. (D) The SEC-MALLS elution profile indicates that the C- and N-terminal endo-proteolytic fragments remain associated in solution, similar to what is observed for the *S. pombe*, *R. norvegicus* and *D. melanogaster* UGGT enzymes (41). (E) Circular Dichroism (CD) of wild-type (wt) *CtUGGT* between 20 °C and 90 °C. No secondary structure changes are observed in recombinant *CtUGGT* by circular dichroism in the interval 20-60 °C. (F) Circular Dichroism melt curve of wt *CtUGGT* and double Cys mutants *CtUGGT*^{N796C/G1118C} and *CtUGGT*^{D611C/G1050C} at $\lambda = 208.5$ nm. This wavelength showed the greatest CD change with temperature. Points were fitted to a Boltzmann sigmoidal fit using GraphPad Prism. Protein Concentrations: wt: 0.1 mg/mL; *CtUGGT*^{N796C/G1118C} mutant: 0.09 mg/mL; *CtUGGT*^{D611C/G1050C}: 0.07 mg/mL. (G) CD spectra of the wt and double Cys *CtUGGT* mutants at T=37 °C.

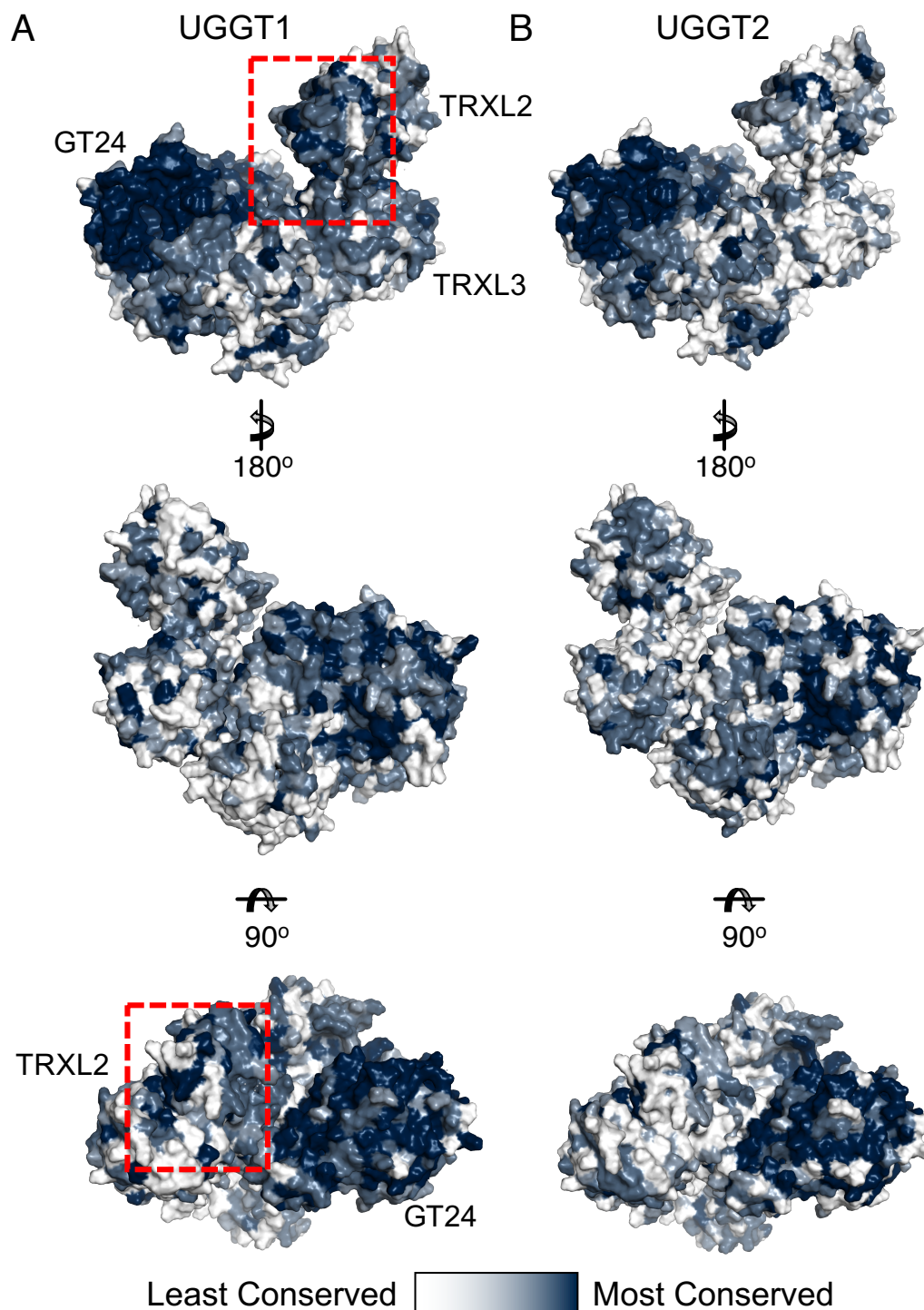


488

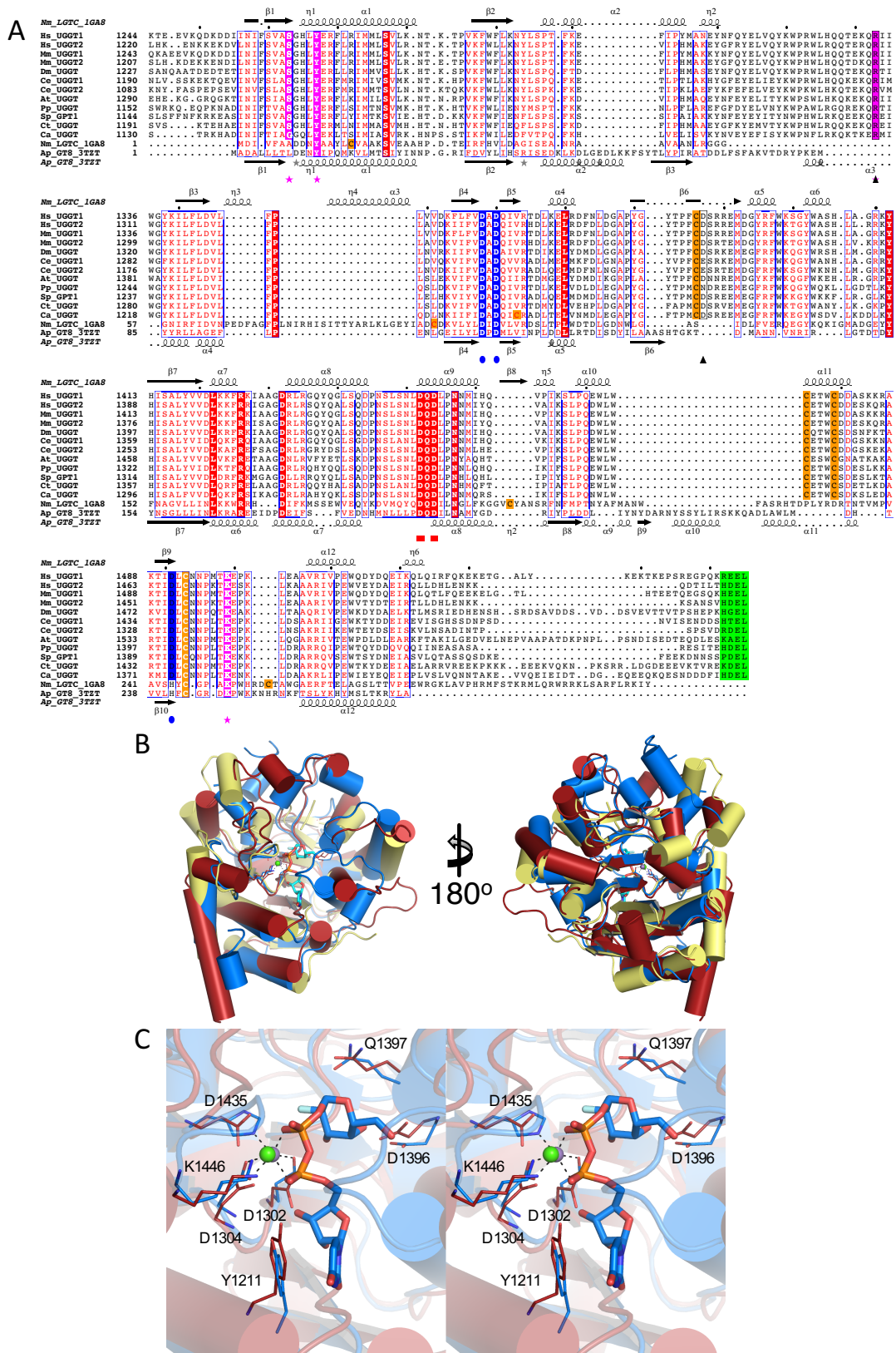
489 **SI Appendix Figure S2. Individual *CtUGGT* domains overlaid on the closest**
 490 **structural homologues in the PDB as per DALI (42) search.** All domain structures
 491 belong to the intermediate form ($P6_1$). (A) *CtUGGT*-TRXL1 in purple, PDB ID 3BD2
 492 in white. Rmsd over 78 C α atoms: 2.0 Å. (B) *CtUGGT*-TRXL2 in blue, PDB ID
 493 3EU3 in white. Rmsd over 132 C α atoms: 1.9 Å. (C) *CtUGGT*-TRXL4 in green, PDB
 494 ID 1UN2 in white. Rmsd over 68 C α atoms: 1.9 Å. (D) *CtUGGT*- β S1 in yellow, PDB
 495 ID 3KPT in white. Rmsd over 72 C α atoms: 1.5 Å. (E) *CtUGGT*- β S2 in orange, PDB
 496 ID 5AQ0 in white. Rmsd over 73 C α atoms: 1.5 Å. (F) *CtUGGT*-GT in red, PDB ID
 497 1GA8 in white. Rmsd over 194 C α atoms: 1.7 Å. The previously determined crystal

498 structure of *Ct*UGGT-TRXL3 (25) has been omitted from this figure. The rmsd
499 between our *Ct*UGGT-TRXL3 and PDB ID 3WZS is 0.9 Å over 142 Cα atoms (0.9
500 Å over 154 Cα atoms for PDB ID 3WZT).

501

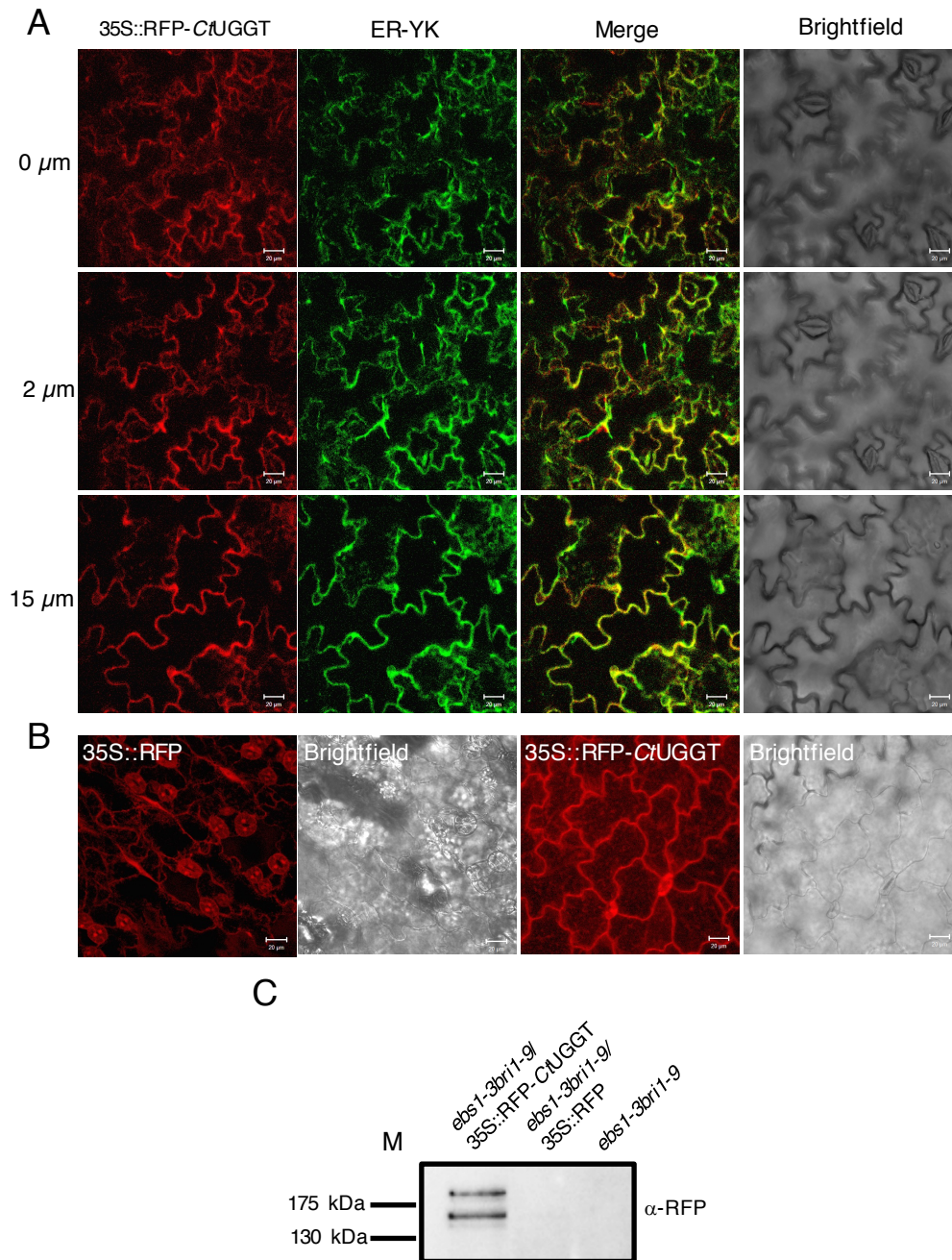


SI Appendix Figure S3. Sequence conservation mapped onto homology models of human UGGT1 and UGGT2. The red boxes mark TRXL2 and TRXL3 surface residues conserved across UGGT1 sequences but not across UGGT2 sequences. These residues are putative misfold-recognition sites. Figure made with the program Consurf (40).



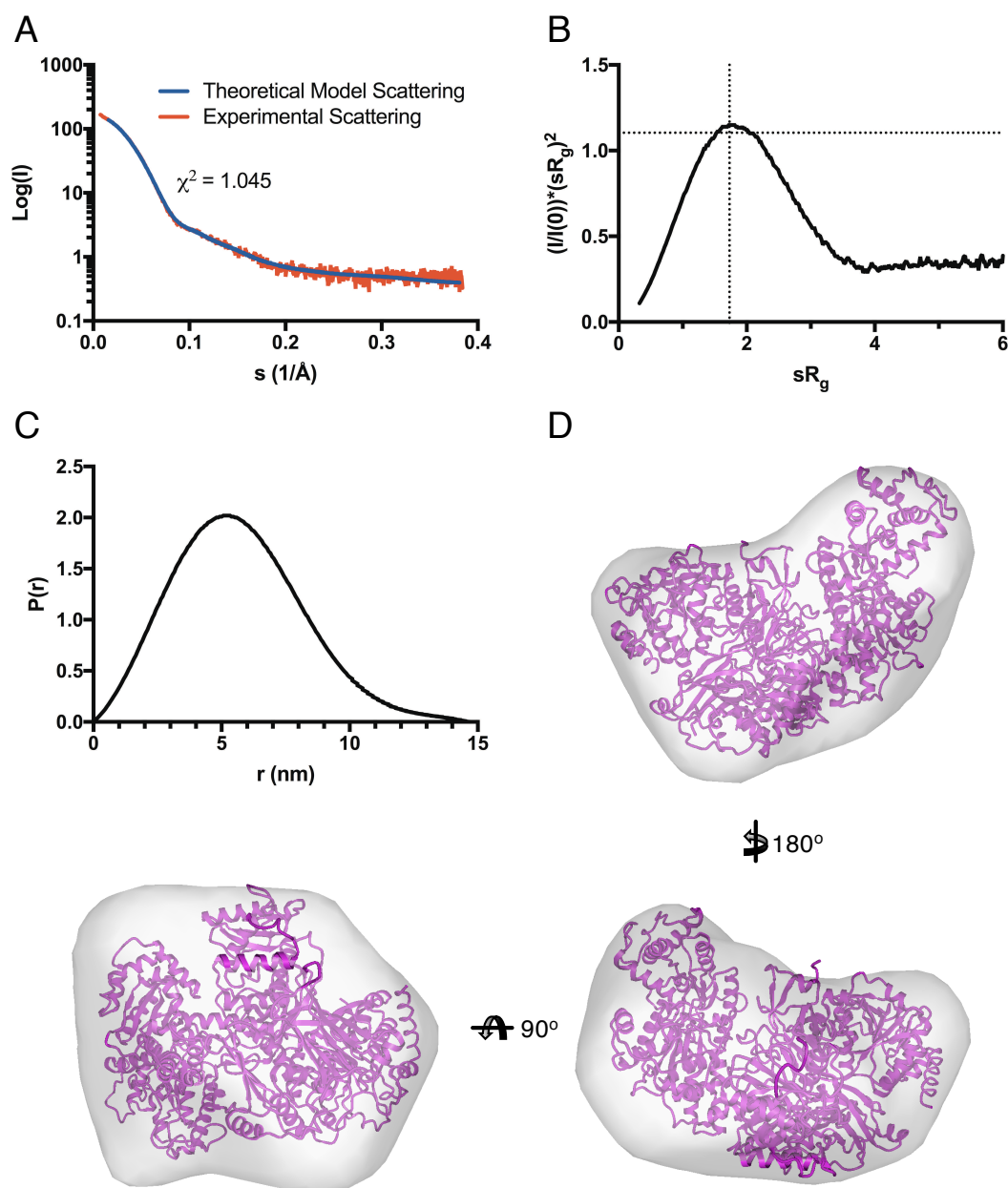
SI Appendix Figure S4. The UGGT catalytic domain (fold family GT24). (A) Sequence alignment of UGGTs from multiple eukaryotic organisms (made using the program Esript (43)). *Hs*: *Homo sapiens*; *Mm*: *Mus musculus*; *Dm*: *Drosophila*

melanogaster; *Ce*: *Caenorhabditis elegans*; *At*: *Arabidopsis thaliana*; *Pp*: *Pichia*
pastoris; *Sp*: *Schizosaccharomyces pombe*; *Ct*: *Chaetomium thermophilum*; *Ca*:
Candida albicans. The figure includes the sequences and secondary structures of two
structurally characterized homologous glycosyltransferases: *Nm*_LGTC_1GA8,
Neisseria meningitidis galactosyltransferase LgtC (PDB ID 1GA8); and
*Ap*_GT8_3TZZ, *Anaerococcus prevotii* glycosyltransferase family 8 (PDB ID 3TZZ).
Red squares below the sequences mark conserved residues that have been implicated
in catalysis, either in UGGT or in homologous glycosyltransferases. Blue squares
mark divalent-metal coordinating residues. Marked by magenta stars, putative UDP-
Glucose coordinating residues. Black triangles mark putative substrate binding
residues. Cys residues are highlighted in gold. **(B)** Structural alignment of the
*Ct*UGGT GT domain (red) with *Neisseria meningitidis* galactosyltransferase LgtC
(PDB ID 1GA8, blue) and *Anaerococcus prevotii* glycosyltransferase family 8 (PDB
ID 3TZZ, yellow). Helices are represented by cylinders. **(C)** Stereo image of the
conserved features for sugar-nucleotide binding, in sticks representation. In red,
*Ct*UGGT; in blue, *Neisseria meningitidis* galactosyltransferase LgtC with 2-Fluoro-2-
deoxy-UDP-Glc bound (PDB ID 1GA8). The uracil moiety in UDP-Glc putatively
stacks on the conserved tyrosine *Ct*UGGT Y1211. The conserved *Ct*UGGT
¹³⁹⁶DQD¹³⁹⁸ motif is the putative binding site for the glucose moiety of UDP-Glc (44).
The manganese (purple) and calcium (green) ions are represented as spheres,
putatively coordinating the UDP-Glc disphosphate and bound by the divalent-metal
coordinating *Ct*UGGT D1435 and the ¹³⁰²DAD¹³⁰⁴ motif (44). The conserved
*Ct*UGGT K1446 may be involved in binding the UDP-Glc disphosphate too.

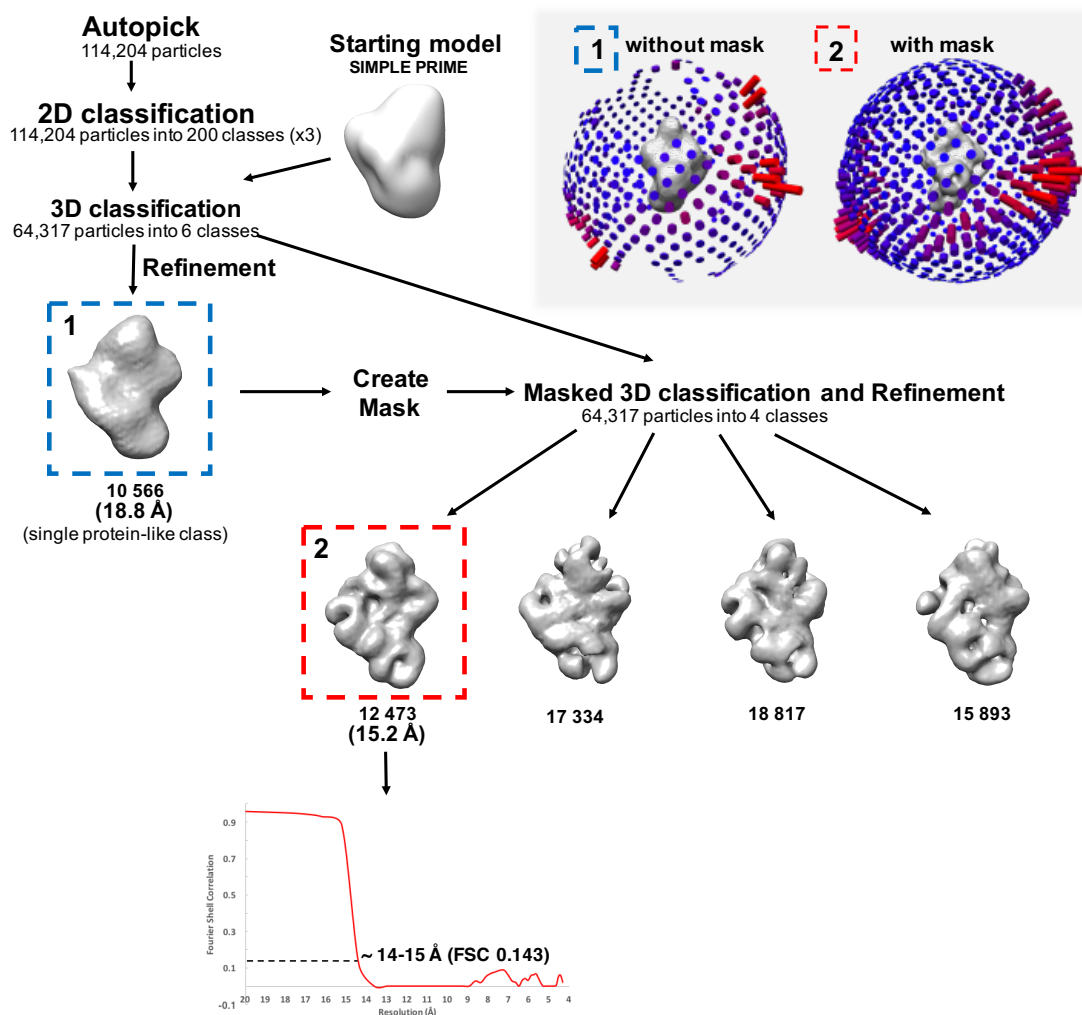


SI Appendix Figure S5. In vivo localisation of *CtUGGT*. (A) *CtUGGT* localizes and folds correctly in the Endoplasmic Reticulum (ER) of *Nicotiana tabacum*. 0 (cortical region), 2 μm and 15 μm confocal images representative of fluorescent proteins imaged by confocal laser scanning microscopy (CLSM) in tobacco leaves (*Nicotiana tabacum*) transiently co-expressing RFP-*CtUGGT* fusion protein under the control of CaMV 35S promoter (35S::RFP-*CtUGGT*) and a yellow-fluorescent ER marker (ER-YK). The reticulated ER network in red co-localizes (merge) with ER-YK. (B) Confocal images of cotyledon epidermal cells of *Arabidopsis thaliana ebs1-3*

bril-9 mutant plants stably expressing 35S::RFP-*CtUGGT*. The network highlighted by the 35S::RFP-*CtUGGT* corresponds to the ER (right panel) and was compared to the one in *ebs1-3 bril-9* mutant plants stably expressing the cytosolic RFP (35S::RFP left panel). Images show the projection of 15 compiled confocal optical sections (1 μ m each) from the cortical to the deepest plane. **(C)** Western blot analysis of microsomal fractions from leaves of *Arabidopsis thaliana ebs1-3 bril-9* mutant and from lines stably expressing 35S::RFP-*CtUGGT* or 35S::RFP. All microsomal fractions were probed with an anti-RFP (a-RFP) antibody. Molecular mass markers (M in kDa) are indicated. The double band is indicative of *CtUGGT* endo-proteolysis in the plant ER, as observed *in vitro* for *CtUGGT* (this work) and for the *S. pombe*, *R. norvegicus* and *D. melanogaster* UGGT enzymes (45, 46).



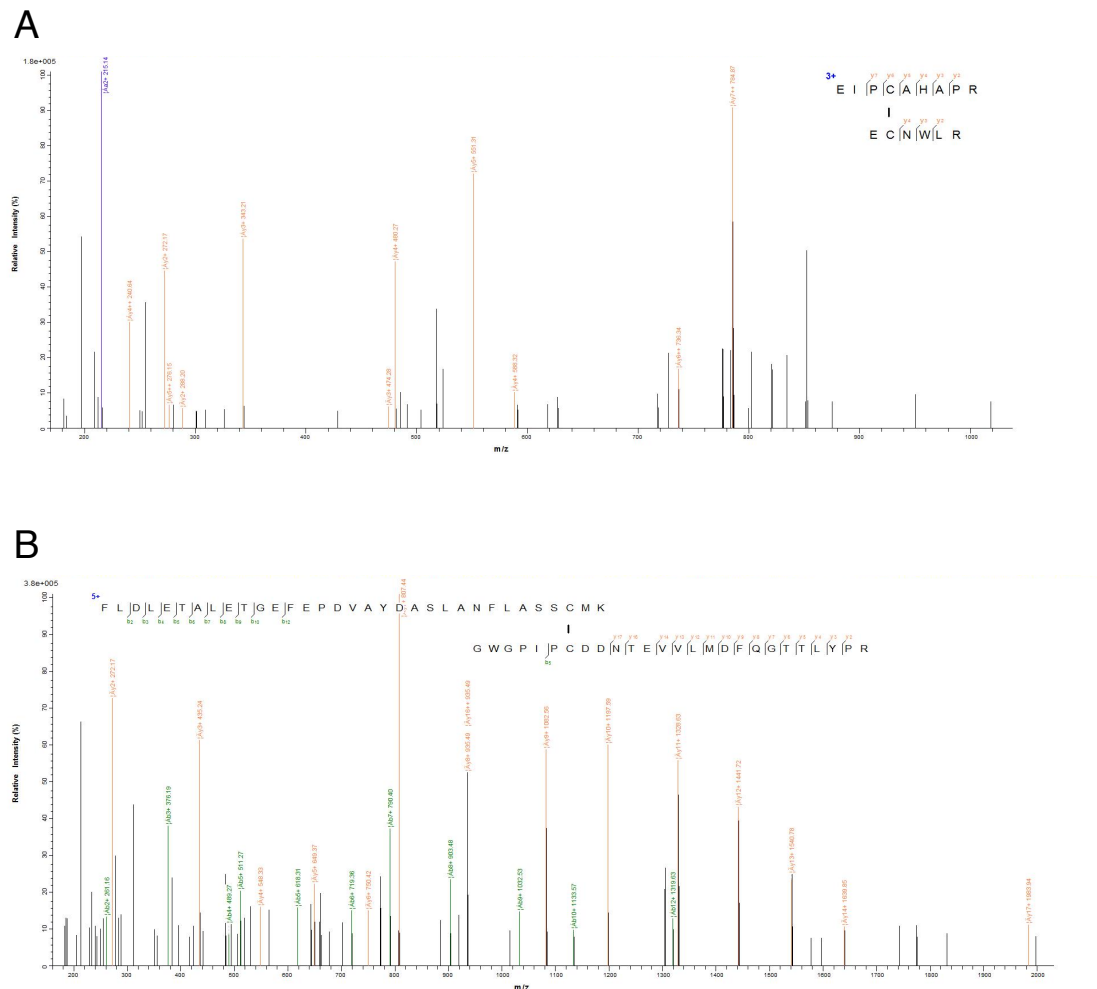
SI Appendix Figure S6. *CtUGGT* SAXS analysis. (A) Experimental scattering curve (red) with associated fit (blue). (B) Normalized Kratky plot for flexibility analysis. (C) Pair distribution function. (D) Three orthogonal views of the *CtUGGT* crystal structure in the open conformation docked in the SAXS envelope filtered at 25 Å.



SI Appendix Figure S7. *CtUGGT* Cryo-EM. Single-particle 3D reconstruction workflow, overview of single-particle 3D classification and refinement from Volta phase plate data, 3D reconstruction and resolution assessment. Initial 2D classification in multiple rounds was used to remove particles contributing to irrelevant ('bad') classes. 3D classification, in which an initial 3D reference produced using the PRIME algorithm, allowed classification of the data into 6 classes. At this stage only a single class appeared protein-like (blue dashed square). Fitting the crystal structures of *CtUGGT* confirmed that this class corresponded to all of *CtUGGT* except the TRXL2 domain. Inspection of the angular distribution plot (grey inset) suggested that resolution was limited by poor particle alignment, with many views in which TRXL2 would contribute to the 'silhouette' of the particle being underrepresented – suggesting continuous flexibility in this region, as 3D classification failed to retrieve a discrete conformation. Classification of the particle stack with masking based on the initial structure resulted in a 15.2 Å map (see the 'Gold-standard' Fourier shell correlation for the final 3D cryo-EM reconstruction

583 consisting of 12,473 particles) that agreed more closely with the crystal structures;
584 this map also revealed defined domains corresponding to all but TRXL2. The
585 necessary masking out of TXRL2 to achieve higher resolution confirms the
586 continuous nature of the flexibility of this domain which negatively impact the
587 alignment of the *Ct*UGGT particles images.
588

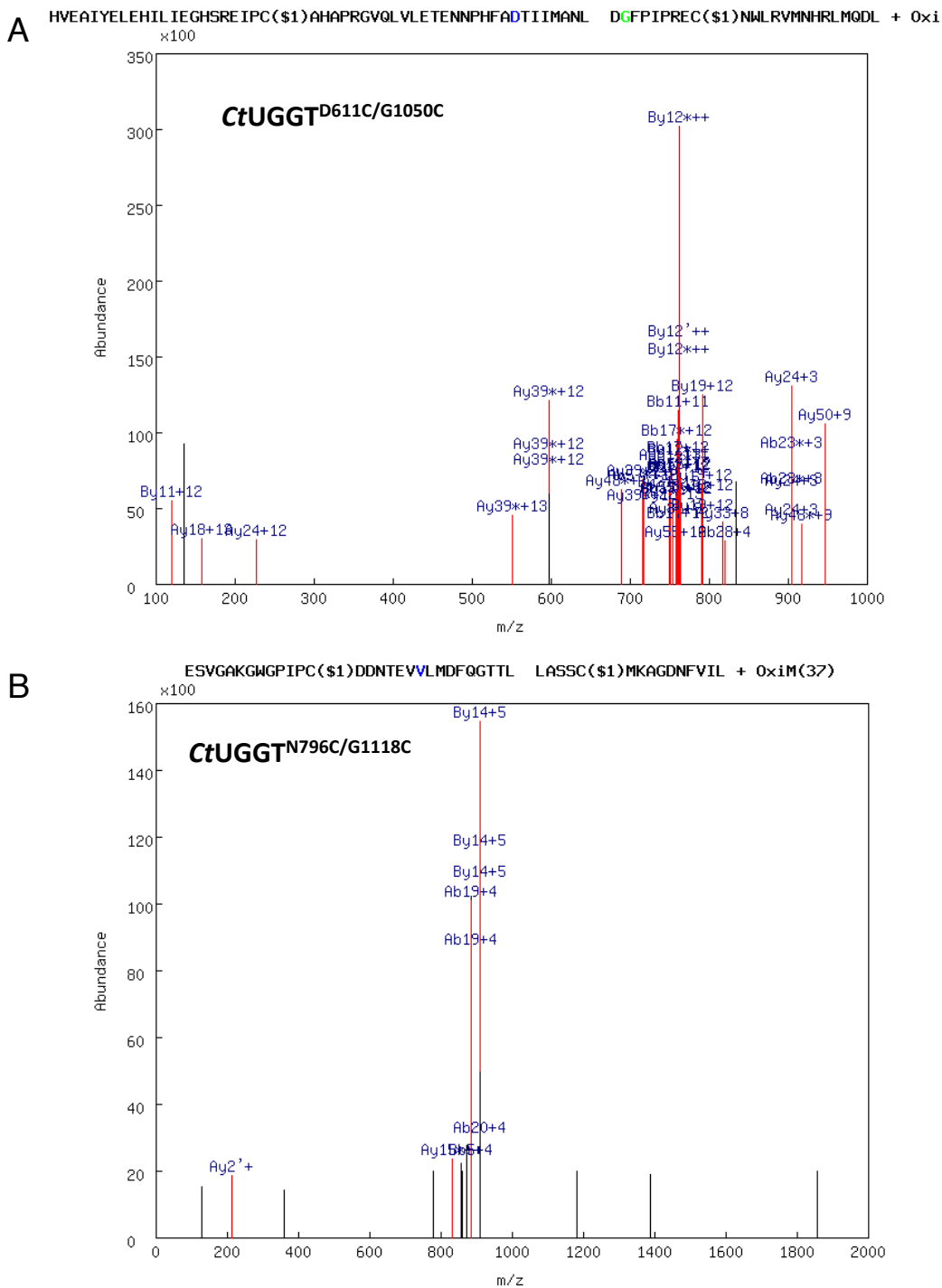
589
590



591

592 **SI Appendix Figure S8. Mass spectrometry of tryptic peptides confirms the**
593 **disulfides in the *Ct*UGGT double Cys mutants.** In peptide mass spectrometry,
594 fragment ions that appear to extend from the amino- or carboxy-terminus of a peptide
595 are termed “b” or “y” ions, respectively. **(A)** Mass spectrometry detection of ions
596 derived from fragmentation of the disulphide-bridged tryptic peptides $^{610}\text{ECNWLR}^{615}$
597 and $^{1047}\text{EIPCAHA}^{1053}$ in the double mutant $\text{CtUGGT}^{\text{D611C/G1050C}}$. The ions confirm the
598 establishment of the engineered disulphide bridge at positions 611-1050 between the
599 TRXL2 and βS2 domains. No peptides containing free Cys at either position 611 or
600 1050 were detected. **(B)** Mass spectrometry detection of ions derived from
601 fragmentation of the disulphide-bridged tryptic peptides
602 $^{766}\text{FLDLETALETGEFEPDVAYDASLANFLASSCMK}^{798}$ and
603 $^{1112}\text{GWGPIPCDDNTEVVLMD FQGTTL YPR}^{1137}$ in the double mutant
604 $\text{CtUGGT}^{\text{N796C/G1118C}}$. The ions confirm the establishment of the engineered disulphide

bridge at positions 796-1118 between the TRXL3 and β S2 domains. No peptides containing free Cys at either position 796 or 1118 were detected.



SI Appendix Figure S9. Mass spectrometry of peptic peptides confirms the disulfides in the *CtUGGT* double Cys mutants. In peptide mass spectrometry, fragment ions that appear to extend from the amino- or carboxy-terminus of a peptide are termed “b” or “y” ions, respectively. **(A)** Mass spectrometry detection of ions derived from fragmentation of the disulphide-bridged peptic peptides A:

613 ¹⁰¹⁷RIKDIKAKRGTEHVEAIYELEHILIEGHSREIPCAHAPRGVQLVLETENNP
 614 FADTIIM[Oxi]ANL¹⁰⁷⁸ and B: ⁶⁰³DGFPIPRECNWLRVM[Oxi]NHRLMQDL⁶²⁵ in
 615 the double mutant CtUGGT^{D611C/G1050C}. The ions confirm the establishment of the
 616 engineered disulphide bridge at positions 611-1050 between the TRXL2 and β S2
 617 domains. No peptides containing free Cys at either position 611 or 1050 were
 618 detected. **(B)** Mass spectrometry detection of ions derived from fragmentation of the
 619 disulphide-bridged peptic peptides A:
 620 ¹¹⁰⁶ESVGAKGWGPIPCDDNTEVVLMDFAQGTTL¹¹³⁴ and B:
 621 ⁷⁹²LASSCM[Oxi]KAGDNFVIL⁸⁰⁶ in the double mutant CtUGGT^{N796C/G1118C}. The
 622 ions confirm the establishment of the engineered disulphide bridge at positions 796-
 623 1118 between the TRXL3 and β S2 domains. No peptides containing free Cys at either
 624 position 796 or 1118 were detected.
 625

626 SI Appendix Tables.

627

628 SI Appendix Table S1. *Ct*UGGT X-ray diffraction data collection statistics.

Structure	WT Closed	WT Intermediate				WT Open	D611C/ G1050C
Space group (<i>Z</i>)	P4 ₃ (8)	P6 ₁ (6)				P6 ₁ 22 (12)	P2 ₁ 2 ₁ 2 ₁ (4)
Wavelength (Å)	0.92821	0.97880	1.07141			0.91741	
Soak	-	-	K ₂ PtCl ₆	K ₂ PtBr ₆	K ₂ PtI ₆	-	-
Cell dimensions <i>a</i> , <i>b</i> , <i>c</i> (Å) <i>α</i> , <i>β</i> , <i>γ</i> (°)	a=b=148.6, c=180.0	a=b=116.9 c=301.7	a=b=117.0 c=305.2	a=b=115.4 c=297.4	a=b=116.9 c=301.5	a=b=163.6 c=248.6	a=78.2, b=142.2, c=186.1
	90,90,90	90,90,120					90,90,90
Resolution range(Å)	180.0-4.3 (4.5-4.3)	301.7-3.8 (4.0-3.8)	152.6-7.0 (7.4-7.0)	99.9-4.4 (4.7-4.4)	150.7-3.5 (3.7-3.5)	124.3-3.5 (3.7-3.5)	77.8-2.8 (2.9-2.8)
<i>R</i> _{merge}	0.7(3.3)	0.3 (3.6)	0.3 (2.3)	0.2 (0.9)	0.2 (3.6)	0.2 (3.1)	0.1 (3.9)
<i>R</i> _{meas}	0.8(3.6)	0.3 (3.9)	0.3 (2.5)	0.2 (1.2)	0.2 (3.7)	0.2 (3.3)	0.2 (4.3)
<i>CC</i> _{1/2}	0.98(0.44)	0.99 (0.37)	0.97 (0.44)	0.98 (0.41)	0.98 (0.48)	0.99 (0.49)	0.99 (0.41)
<i>I</i> / <i>σI</i>	4.9 (1.3)	9.7 (0.8)	8.9 (1.2)	6.4 (1.2)	15.4 (1.3)	9.2 (0.8)	7.8 (0.6)
Completeness (%)	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)	99.3 (99.9)	96.2 (100.0)	100.0 (100.0)	99.7 (99.8)
Redundancy	13.9(14.4)	11.7(11.7)	11.4(11.7)	3.9(3.9)	23.2(23.3)	11.8(12.2)	5.0 (5.2)

629

630 **SI Appendix Table S2. *Ct*UGGT P6_I – Intermediate. MIR-AS phasing statistics.**
 631 The fluorescence measurements at the Pt L-III edge gave $f'_{\text{Pt}}=-27.7 \text{ e}^-$ and $f''_{\text{Pt}}=14.18$
 632 e^- . These values were kept fixed during SHARP heavy-atom refinement.

	Native	K ₂ PtCl ₆	K ₂ PtBr ₆	K ₂ PtI ₆
Data collection		BM14@ESRF		
# Pt sites	-	5	7	6
Average Pt occupancy	-	0.38	0.49	0.35
Pt Bfactor (Å²)		167	167	127
PhasingPower* (ano, acentrics)	-	0.19 (15.0Å)	0.19 (10.7Å)	0.47 (6.2Å)
Phasing Power* (iso, acentrics)	-	0.57 (15.0Å)	0.40 (15.0Å)	- [§]
Phasing Power* (iso,centrics)	-	0.51 (15.0Å)	0.40 (15.0Å)	- [§]
	0.93 (6.2Å)			
FOM (centrics)	0.23 (0.21 in the 6.78-6.19 Å resolution shell)			
FOM (acentrics)	0.24 (0.44 in the 6.78-6.19 Å resolution shell)			

633 [§]: the K₂PtI₆ was used as a reference dataset because of its highest resolution and in
 634 order to avoid correlated non-isomorphism; as a result, no isomorphous differences
 635 phasing power statistics were computed for this dataset.

636 *: in parentheses, the resolution past which the phasing power falls below 1.0

637 **SI Appendix Table S3. *Ct*UGGT crystal structures, refinement statistics.**

Crystal form (Z)	P4 ₃ (8) (WT closed)	P6 ₁ K ₂ PtI ₆ (6) (WT intermediate)	P6 ₁ 22 (12) (WT open)	P2 ₁ 2 ₁ 2 ₁ (4) (D611C/G1050C)
PDB ID	5N2J	5MU1	5MZO	5NV4
Resolution (Å)	114.6-4.3 (4.5-4.3)	101.3 (3.5) (3.6-3.5)	141.7-3.5 (3.6-3.5)	77.9-2.8 (2.9-2.8)
No. reflections	20988 (1304)	25903 (1269)	25379 (2359)	51917 (3033)
R_{work} / R_{free}	0.201/0.259 (0.229/0.275)	0.237/0.252 (0.230/0.270)	0.251/0.278 (0.253/0.306)	0.239/0.245 (0.299/0.304)
No. atoms	22527	11720	11227	11198
Protein	22525	11411	11035	11066
Ligand/ion	2 Ca ²⁺	6 Pt ⁴⁺ , 14 I, 1 Ca ²⁺	1 Ca ²⁺	-
Water	0	0	12	34
B-factors (Å ²)				
Protein	160	149	170	110
Ligand/ion	218	243	167	-
Water	-	-	79	71
R.m.s. deviations				
Bond lengths (Å)	0.01	0.01	0.01	0.01
Bond angles (°)	1.0	0.24	1.2	1.1

638 All structures were refined against X-ray data from one crystal only. Values in
639 parentheses are for highest-resolution shell.

640

641

642 **SI Appendix Table S4. SAXS data and modelling statistics.**

643

Data-Collection parameters

BeamLine	BM29, ESRF , Grenoble France
Detector	Pilatus 1M
Beam size	0.7 x 0.7 mm
Energy	12.5 keV
Sample-to-detector distance (mm)	2867
Capillary diameter (mm)	1.8
q range (\AA^{-1})	0.0025 - 0.5
Exposure time (s)	10 x 1
Temperature (K)	293
Concentration range (mg mL^{-1})	0.06 – 2.4

Structural parameters

I(0) BSA	74
q Interval for Fourier inversion (\AA^{-1})	0.016 – 0.16
I(0) [from P(r)]	160.83 ± 0.22
R _g [from P(r)] (\AA)	43.7 ± 0.7
I(0) [from Guiner approximation]	160.90 ± 0.22
R _g [from Guiner approximation] (\AA)	43.5 ± 0.8
sR _g limits [from Guiner approximation]	0.69 - 1.26
D _{max} (\AA)	147
Porod volume estimate (nm^3)	360
DAMMIF excluded volume (nm^3)	362
Molecular Mass (kDa)	
From Porod (x 0.53 - 0.63)	185 - 225
From excluded volume (x 0.5)	182
From I(0)	143
From sequence	174

Software employed

Primary data reduction	PIPELINE
Data processing	PRIMUS /ScÅtter
<i>Ab initio</i> modelling	DAMMIF/DAMMIN
Validation and averaging	DAMAVER/DAMCLUST
Computation of model intensities	CRY SOL

644

645 $q = 4\pi\sin(\theta/\lambda)$, where 2θ is the scattering angle and λ is the wavelength

646

SI Appendix Table S5. DNA primers used in this work

Prim er	Sequence
1F	5'-ctaaagaagaaggggtatctctcgagaaaagaCAAGTCGCAGCCTCTCCA-3'
1R	5'-ctgagatgagttttgttctagaccTCCCGAACCGTCTTGAC-3' (*)
2F	5'-GACTGGTTCCAATTGACAAGC-3'
2R	5'-GCAAATGGCATTCTGACATCC-3'
3F	5'-GAAACAGCAGCGTCCGAC-3'
3R	5'-CAGCAGCTCGACCAAGTATG-3'
4F	5'-GGTTGCGTAGCTGAAACCGGTCAAGTCGCAGCCTCTCCA-3'
4R	5'-GATGGTGGTGCTTGGTACCCTCCCGAACCGTCTTGAC-3'
5F	5'-CTACAGCTCCTGGGCAACGTG-3'
5R	5'-CATTGGCCACACCAGCCAC-3'
6F	5'- GGGGACAAGTTTGTACAAAAAAGCAGGCTTTGTTCGCAGCCTCTCC ATCAA-3'
6R	5'- GGGGACCACTTTGTACAAGAAAGCTGGGTCTCACAGCTCATCCTTC TCCCGAACCGTCTTGAC-3'
7F	5'-GCCCTATTCCTtGCGATGACAACACTG-3'
7R	5'-CAGTGTTGTCATCGCaAGGAATAGGGC-3'
8F	5'-TTGCCAGCTCAtgCATGAAAGCTGGC-3'
8R	5'-GCCAGCTTTCATGcaTGAGCTGGCAA-3'
9F	5'-GCGAGATTCCCtGcGCTCATGCCCCAC-3'
9R	5'-GTGGGGCATGAGCgCaGGGAATCTCGC-3'
10F	5'-TTCCACGGGAGtgCAATTGGCTGCGGG-3'
10R	5'-CCCGCAGCCAATTGcaCTCCCGTGGAA-3'

(*) the bases coding for the C-terminal ¹⁵⁰⁶KDEL¹⁵⁰⁹ *Ct*UGGT ER-localization signal were omitted from this reverse primer.

SI Appendix Movies S1. Two orthogonal views of the morphing between three conformations of wt CtUGGT. (A,B) The transition depicted is from the “open” conformation (P6₁22) to the “intermediate” one (P6₁) to the “closed” one (P4₃). The protein is in cartoon representation, its seven domains colored as detailed for Figure 1. The structures were first aligned using Theseus (47) and the morphing calculated in PyMOL.

References

1. Amlacher S, et al. (2011) Insight into Structure and Assembly of the Nuclear Pore Complex by Utilizing the Genome of a Eukaryotic Thermophile. *Cell* 146(2):277–289.
2. Aricescu AR, Lu W, Jones EY (2006) A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta Crystallogr D Biol Crystallogr* 62(Pt 10):1243–1250.
3. Whitmore L, Wallace BA (2008) Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers* 89(5):392–400.
4. Trombetta SE, Bosch M, Parodi AJ (1989) Glucosylation of glycoproteins by mammalian, plant, fungal, and trypanosomatid protozoa microsomal membranes. *Biochemistry* 28(20):8108–8116.
5. Caputo AT, et al. (2016) Structures of mammalian ER α -glucosidase II capture the binding modes of broad-spectrum iminosugar antivirals. *Proc Natl Acad Sci USA* 113(32):E4630–8.
6. Petoukhov MV, et al. (2012) New developments in the ATSAS program package for small-angle scattering data analysis. *J Appl Crystallogr* 45(2):342–350.
7. la Rosa-Trevín de JM, et al. (2016) Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy. *J Struct Biol* 195(1):93–99.
8. Rohou A, Grigorieff N (2015) CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* 192(2):216–221.
9. Elmlund H, Elmlund D, Bengio S (2013) PRIME: probabilistic initial 3D model generation for single-particle cryo-electron microscopy. *Structure* 21(8):1299–1306.
10. Elmlund D, Elmlund H (2012) SIMPLE: Software for ab initio reconstruction of heterogeneous single-particles. *J Struct Biol* 180(3):420–427.
11. Pettersen EF, et al. (2004) UCSF Chimera--a visualization system for

688 exploratory research and analysis. *J Comput Chem* 25(13):1605–1612.

689 12. Chambers MC, et al. (2012) A cross-platform toolkit for mass spectrometry
690 and proteomics. *Nature Biotechnology* 30(10):918–920.

691 13. Yang B, et al. (2012) Identification of cross-linked peptides from complex
692 samples. *Nat Methods* 9(9):904–906.

693 14. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates,
694 individualized p.p.b.-range mass accuracies and proteome-wide protein
695 quantification. *Nature Biotechnology* 26(12):1367–1372.

696 15. Adusumilli R, Mallick P (2017) Data Conversion with ProteoWizard
697 msConvert. *Methods Mol Biol* 1550(10):339–368.

698 16. Xu H, Freitas MA (2009) MassMatrix: a database search program for rapid
699 characterization of proteins and peptides from tandem mass spectrometry data.
700 *Proteomics* 9(6):1548–1555.

701 17. Gorrec F (2015) The MORPHEUS II protein crystallization screen. *Acta*
702 *Crystallogr F Struct Biol Commun* 71(Pt 7):831–837.

703 18. Evans G, Pettifer RF, IUCr (2001) CHOOCH: a program for deriving
704 anomalous-scattering factors from X-ray fluorescence spectra. *J Appl*
705 *Crystallogr* 34(1):82–86.

706 19. Kabsch W (2010) XDS. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):125–132.

707 20. Evans PR, Murshudov GN (2013) How good are my data and what is the
708 resolution? *Acta Crystallogr D Biol Crystallogr* 69(Pt 7):1204–1214.

709 21. Sheldrick GM, Schneider TR (2002) *Substructure solution with SHELXD* (Acta
710 *Crystallogr. D*).

711 22. Vonrhein C, Blanc E, Roversi P, Bricogne G (2005) *Automated structure*
712 *solution with autoSHARP*. In “*Crystallographic Methods*,” S Doublie, Ed.

713 23. La Fortelle de E, Bricogne G (1997) SHARP: A maximum-likelihood heavy-
714 atom parameter refinement program for the MIR and MAD methods. *Methods*
715 *Enzymol*.

716 24. Vagin A, Teplyakov A (2010) Molecular replacement with MOLREP. *Acta*
717 *Crystallogr D Biol Crystallogr* 66(Pt 1):22–25.

718 25. Zhu T, Satoh T, Kato K (2014) Structural insight into substrate recognition by
719 the endoplasmic reticulum folding-sensor enzyme: crystal structure of third
720 thioredoxin-like domain of UDP-glucose:glycoprotein glucosyltransferase. *Sci*
721 *Rep* 4:7322.

722 26. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development
723 of Coot. *Acta Crystallogr D Biol Crystallogr* 66(Pt 4):486–501.

- 724 27. Bricogne G, Blanc E, Brandl M, Flensburg C, Keller P (2011) *autoBUSTER*.
725 *Cambridge United Kingdom* (Global Phasing Ltd).
- 726 28. Schröder GF, Levitt M, Brunger AT (2014) Deformable elastic network
727 refinement for low-resolution macromolecular crystallography. *Acta*
728 *Crystallogr D Biol Crystallogr* 70(Pt 9):2241–2255.
- 729 29. DiMaio F, et al. (2011) Improved molecular replacement by density- and
730 energy-guided protein structure optimization. *Nature* 473(7348):540–543.
- 731 30. Adams PD, et al. (2010) PHENIX: a comprehensive Python-based system for
732 macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt
733 2):213–221.
- 734 31. Smart OS, et al. (2012) Exploiting structure similarity in refinement: automated
735 NCS and target-structure restraints in BUSTER. *Acta Crystallogr D Biol*
736 *Crystallogr* 68(Pt 4):368–380.
- 737 32. McCoy AJ, et al. (2007) Phaser crystallographic software. *J Appl Crystallogr*
738 40(Pt 4):658–674.
- 739 33. Clough SJ, Bent AF (1998) Floral dip: a simplified method for *Agrobacterium*-
740 mediated transformation of *Arabidopsis thaliana*. *Plant J* 16(6):735–743.
- 741 34. Murashige T, Skoog F (1962) A Revised Medium for Rapid Growth and Bio
742 Assays with Tobacco Tissue Cultures. *Physiologia Plantarum* 15(3):473–497.
- 743 35. Nelson BK, Cai X, Nebenführ A (2007) A multicolored set of in vivo organelle
744 markers for co-localization studies in *Arabidopsis* and other plants. *Plant J*
745 51(6):1126–1136.
- 746 36. Batoko H, Zheng HQ, Hawes C, Moore I (2000) A rab1 GTPase is required for
747 transport between the endoplasmic reticulum and golgi apparatus and for
748 normal golgi movement in plants. *Plant Cell* 12(11):2201–2218.
- 749 37. Brandizzi F, et al. (2002) The destination for single-pass membrane proteins is
750 influenced markedly by the length of the hydrophobic domain. *Plant Cell*
751 14(5):1077–1092.
- 752 38. Avila JR, Lee JS, Torii KU (2015) Co-Immunoprecipitation of Membrane-
753 Bound Receptors. *Arabidopsis Book* 13:e0180.
- 754 39. Webb B, Sali A (2016) Comparative Protein Structure Modeling Using
755 MODELLER. *Curr Protoc Protein Sci* 86(Suppl 1):2.9.1–2.9.37.
- 756 40. Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N (2010) ConSurf 2010:
757 calculating evolutionary conservation in sequence and structure of proteins and
758 nucleic acids. *Nucleic Acids Res* 38(Web Server issue):W529–33.
- 759 41. Albesa-Jové D, Guerin ME (2016) The conformational plasticity of
760 glycosyltransferases. *Curr Opin Struct Biol* 40:23–32.

- 761 42. Holm L, Rosenström P (2010) Dali server: conservation mapping in 3D.
762 *Nucleic Acids Res* 38(Web Server issue):W545–9.
- 763 43. Robert X, Gouet P (2014) Deciphering key features in protein structures with
764 the new ENDscript server. *Nucleic Acids Res* 42(Web Server issue):W320–4.
- 765 44. Arnold SM, Fessler LI, Fessler JH, Kaufman RJ (2000) Two Homologues
766 Encoding Human UDP-Glucose:Glycoprotein Glucosyltransferase Differ in
767 mRNA Expression and Enzymatic Activity †. *Biochemistry* 39(9):2149–2163.
- 768 45. Parker CG, Fessler LI, Nelson RE, Fessler JH (1995) Drosophila UDP-
769 glucose:glycoprotein glucosyltransferase: sequence and characterization of an
770 enzyme that distinguishes between denatured and native proteins. *The EMBO*
771 *Journal* 14(7):1294–1303.
- 772 46. Guerin M, Parodi AJ (2003) The UDP-glucose:glycoprotein
773 glucosyltransferase is organized in at least two tightly bound domains from
774 yeast to mammals. *Journal of Biological Chemistry* 278(23):20540–20546.
- 775 47. Theobald DL, Steindel PA (2012) Optimal simultaneous superpositioning of
776 multiple structures with missing data. *Bioinformatics* 28(15):1972–1979.
- 777