**OS R PR-00005 15.01.2020**
**Crystal structure determination of *Ct*UGGT$_{Kif}$ from HEK293F cells treated with 5 µM kifunensine.**

**Summary.** *Ct*UGGT$_{Kif}$ was purified from the supernatant of HEK293F cells treated with 5 mM kifunensine and transfected with the pHLsec:*Ct*UGGT vector for secreted expression of *Chaetomium thermophilum* UGGT (*Ct*UGGT). The protein carries high-mannose glycans, as evidenced by EndoF1 cleavage. A crystal was grown from the *Ct*UGGT$_{Kif}$ sample and diffraction data were collected on I03@Diamond on 01.05.2016. Structure factor amplitudes scaled to 4.1 Å resolution (after anisotropic scaling) and the phases were determined by molecular replacement using PDB ID 5NV4 as a search model. The crystals belong to the same $P2_12_12_1$ crystal form already deposited as PDB ID 5NV4 for the *Ct*UGGT D611C:G1050C double mutant. The *Ct*UGGT$_{Kif}$ structure (deposited as PDB ID 6TRF) combines the close inter-domain distance between the TRXL2 and βS2 domains observed in the *Ct*UGGT "closed-like conformation" (PDB ID 5NV4) with the larger inter-domain distance between the TRXL1 and TRXL3 domains observed in the *Ct*UGGT "open conformation" (PDB ID 5MZO). This *Ct*UGGT$_{Kif}$ crystal structure suggests that the UGGT molecule can close the gap between the TRXL2 and βS2 domains while at the same time opening the one between the TRXL1 and TRXL3 domains. The disorder observed in the TRXL2 domain in the crystal may be a clue that UGGT bears an intrinsically disordered domain, a finding that would explain how UGGT can recognise misfolded glycoproteins.

***Ct*UGGT$_{Kif}$ Protein expression.**
All reagents for transfection were supplied by ThermoFisher Scientific. To express *Ct*UGGT$_{Kif}$, 300 mL human epithelial kidney FreeStyle 293 (HEK293F) cells at $10^6$ cells/mL, suspended in GIBCO FreeStyle 293 Media, supplemented with 5 µM kifunensine (Cayman Chemical Company), were transfected using the FreeStyle MAX 293 expression system, according to manufacturer instructions (see also (1)). >90 % cell viability was confirmed by trypan blue exclusion. A volume of 375 µL of FreeStyle MAX transfection reagent and 300 µg plasmid DNA in 375 µL of water were each separately diluted to 6 mL with OptiPRO SFM reagent, then mixed and incubated for 7 minutes at room temperature. The mixture was split evenly between two cultures, each containing total 150 mL of HEK293F cells at a density of $10^6$ cells/mL. Transfected cells were left shaking at 135 revolutions per minute in 0.5 L Erlenmeyer flasks with 0.2 µm vent caps (Corning), in an incubator at 37 °C, with 5 % $CO_2$ present, for 6 days.

***Ct*UGGT$_{Kif}$ Protein purification.**
The HEK293F cells were separated from the supernatant by centrifugation for 15 minutes at 4 °C and 3,000 g. The supernatant was made 1x phosphate buffered saline (PBS) and 5 mM imidazole by adding appropriate stock solution volumes, and the pH adjusted to 7.4 by adding a few drops of 2 M NaOH, before vacuum filtration through a 0.45 µm filter and application onto

a 1 mL HisTrap HP Ni IMAC column (GE Healthcare) equilibrated against binding buffer: 1x PBS, 5 mM imidazole, pH adjusted to 7.4 with a few drops of 2 M NaOH. The column was washed with 20 column volumes (cV) buffer A and bound proteins eluted with a linear gradient over 20 cV from 0 % to 100 % of elution buffer B: 1x PBS, 400 mM imidazole, pH adjusted to 7.4 with 2 M NaOH. Peak fractions were pooled and concentrated using a PES membrane, 50 kDa MW cutoff centrifugal ultrafiltration device (Sartorius), to a volume of 5 mL. Concentrated *Ct*UGGT$_{Kif}$ sample was applied to a HiLoad Superdex 200 16/60 size exclusion chromatography (SEC) column (GE Healthcare) equilibrated against SEC buffer: 20 mM NaHEPES, 150 mM NaCl. Peak fractions were pooled and concentrated as before, protein concentration measured by loading 1.5 µL of sample on a NanoDrop 1000 spectrophotometer (Thermo Scientific). The calculated $\varepsilon_{280}$ of *Ct*UGGT$_{Kif}$ = 1.13 mg mL$^{-1}$cm$^{-1}$ was then used to estimate the protein concentration. The final concentration of protein (1 mL volume) was 7.24 mg/mL ($A_{280}$ = 8.18). *Ct*UGGT$_{Kif}$ protein aliquots were frozen in liquid $N_2$ and stored at -80 °C. SDS-PAGE of SEC fractions was used to assess purity. All chromatography was at 1 mL/min flow rate on ÄKTA Pure (room temperature) or ÄKTA Start (4 °C) systems (GE Healthcare).

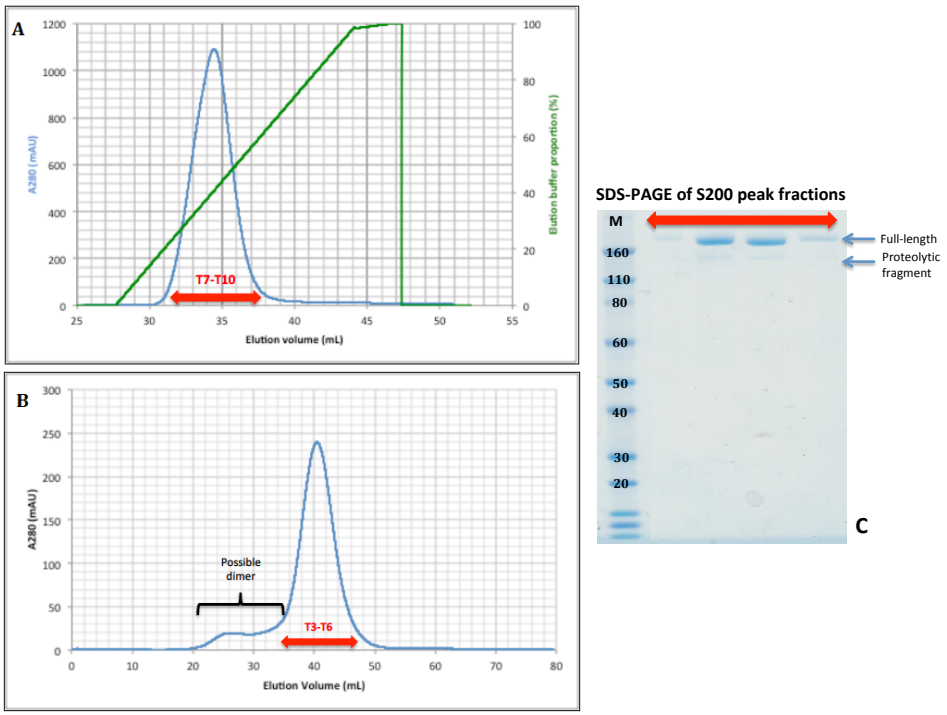


**Figure 1. *Ct*UGGT-6xHis purification. (A):** IMAC elution. **(B)** SEC elution. Peak fractions that were pooled and concentrated are indicated in red. $A_{\infty}$ (mAu)= blue line. Elution buffer percentage: green line. **(C)**. SDS-PAGE analysis of S200 fractions. 1 µL of fraction was diluted in 14 µL milliQ $H_2O$ and 5 µL loading dye, and 20 µL samples loaded on a SDS-PAGE 4-12% NuPAGE gel (Invitrogen). M = MW markers lane. Lanes 2,3,4,5: T3-T6 S200 elution fractions.

**Analysis of the *Ct*UGGT$_{Kif}$ N-linked glycans.**

To confirm that kifunensine inhibited Golgi glycan processing, and that the protein carries high-mannose glycans, such as $GlcNAc_2Man_9$, two sequential digestions of *Ct*UGGT$_{Kif}$ with EndoF1 and PNGaseF were carried out (the former enzyme is unable to cleave complex glycans of the kind processed in the Golgi in absence of mannosidase inhibitors, see Figure 2A).
Glycosidase enzyme:substrate ratios were based on (2). 0.29 mg of *Ct*UGGT$_{Kif}$ was diluted with and dialysed against SEC buffer to remove EDTA from the solution. Dialysis was performed in a 10 kDa MW-cutoff slide-A-lyzer dialysis unit (TFS), against 40 mL SEC buffer at 4 °C. The buffer was changed once between two 3-hour incubations.

For 2AA-labeling of glycans, purification of 2AA-labeled glycans and cleavage and detection of glycans by normal-phase high-performance liquid chromatography, see methods in (3).

**Endoglycosidase F1 (EndoF1) digest**: 6.6 µg of a glutathione-S-transferase (GST) fusion of EndoF1 (1.0 mg/mL, a gift from from Dr Elena Seiradake) was added to the *Ct*UGGT$_{Kif}$ sample and to a *Ct*UGGT control, 0.29 mg each, corresponding to a 1:15 molar ratio of EndoF1-GST:*Ct*UGGT; samples were then incubated at 30 °C overnight. Each sample was then split into two halves: one half was heat inactivated at 95 °C for 3 minutes and stored at -20 °C, while the other half was injected onto a Superdex 200 Increase 10/300 GL column (GE healthcare) equilibrated against SEC buffer, on an ÄKTA Pure system. Peak fractions were pooled and concentrated in PES membrane 50 kDa MW-cutoff centrifugal ultrafiltration devices (Sartorius), to a volume of 0.5 mL.

**Peptide-N-glycosidase F (PNGase F) digest**: 10x New England Biolabs (NEB) glycoprotein denaturing buffer was added to each 0.5 mL sample to 1x concentration, and samples heated at 95 °C for 10 minutes to denature the glycoprotein, then put on ice. 10 % NP-40 (NEB) was added to a 1 % concentration, and 10x NEB glycobuffer 2 to 1x concentration. PNGase F (1.0 mg/mL) was added to give a 1:25 molar ratio of PNGase F:*Ct*UGGT in each sample, and samples incubated at 30 °C overnight, before heat inactivation at 75 °C for 10 minutes, and storage at -20 °C.

Figure 3 shows the HPLC elution profile of 2AA-labelled N-linked glycans stripped from *Ct*UGGT$_{Kif}$ (a1) and *Ct*UGGT after sequential digestions with PNGaseF (a2) and Endo-F1 (b1). a1: Glycans after Endo-F1 treatment of *Ct*UGGT: no peaks (all glycans are complex ones). a2: Glycans after PNGaseF treatment of the Endo-F1 treated *Ct*UGGT: complex glycans. b1: *Ct*UGGT$_{Kif}$ carries N-linked glycans that EndoF1 is able to cleave, therefore they must be high-mannose glycans; b2: glycans after PNGaseF treatment of the Endo-F1 treated *Ct*UGGT$_{Kif}$: as Endo-F removed the high-mannose glycans, no glycans are left.

**Crystal growth and cryoprotection**

The orthorhombic $P2_12_12_1$ *Ct*UGGT$_{Kif}$ crystal comes from drop C10 of Tray 126, bar code label MC001676. The condition is C10 of the MORPHEUS screen [0.09 M NPS, *aka* 0.03M Sodium nitrate, 0.03 Sodium phosphate dibasic, 0.03M Ammonium sulfate; 0.1M Buffer System 3, aka Tris Bicine pH 8.5; 8.530% v/v Precipitant Mix 2, aka 40% v/v Glycerol, 20% w/v PEG 4000] mixed in protein:mother liquor ratio 100 nL:100 nL. The crystal grew at 18 °C and it was cooled down in liquid $N_2$ - self-cryoprotected as it comes from a MORPHEUS screen.

**X-ray diffraction.**

X-ray diffraction from the crystal *Ct*UGGT T126-C10 was measured at I03@Diamond, images /dls/i03/data/2016/mx12346-29/Zitzmann/CtUGGT/T126-C10/T126-C10_2,3,4,5_####.cbf. Data collection parameters: all four batches collected at λ=0.97630 Å, beam size 80x20μm, 0.2° oscillation. Recentring followed after each exposure. Batches 2,3: plate set at 2.9 Å max resolution; batches 4,5, plate set at 3.5 Å max resolution. Batch 2: 450 images, 0.1 s exposure, 70% T. Batches 3,4: 500 images, 0.2 s exposure, 100% T. Batch 5: 350 images, 0.5 s exposure, 100% T.

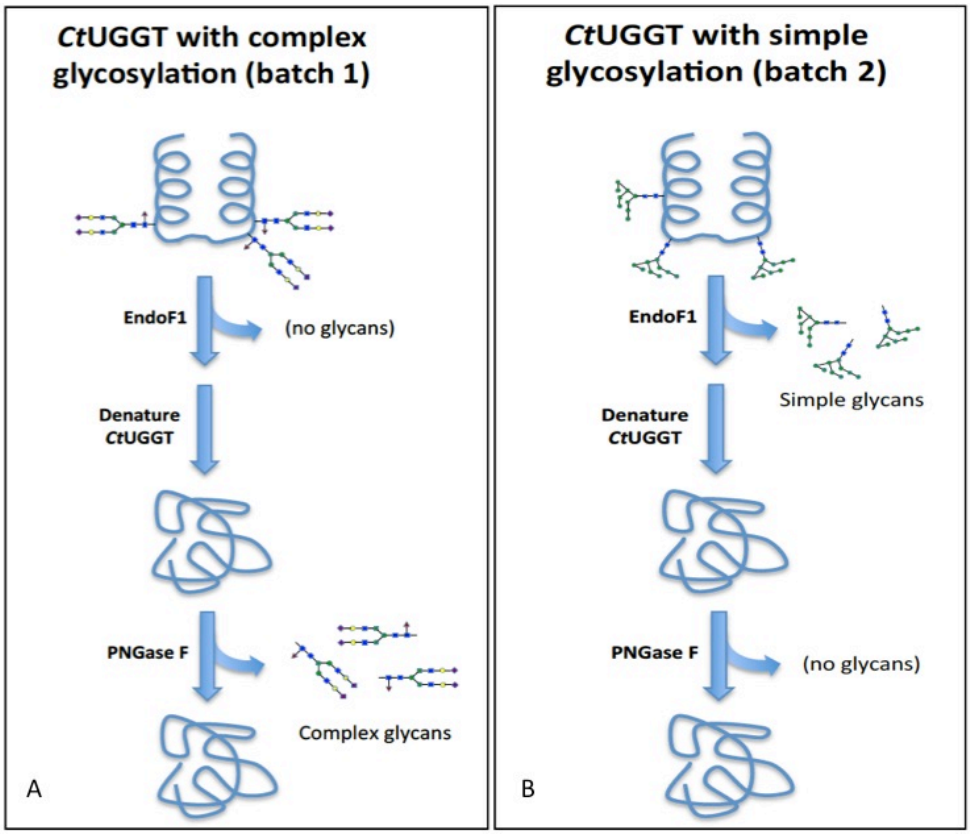


**Figure 2. Sequential digestions with EndoF1 and PNGaseF. A.** Protein expressed without of kifunensine carries N-linked glycans that EndoF1 is unable to cleave **B.** Protein expressed in presence of kifunensine carries N-linked glycans that EndoF1 is able to cleave.

## Data processing

X-ray diffraction datasets were processed with the autoPROC suite of programs (4). Table 1 contains the data processing statistics. CCP4-Matthews was run in $P2_12_12_1$, declaring a molecule of MW~174,000 Dalton. No extra NCS twofolds are visible in the Self Rotation Function section (not shown) with κ=180° (only crystallographic axes are visible), so it is highly likely that the crystal contains 1 copy of *Ct*UGGT$_{Kif}$/asu and 62% solvent.

## *Ct*UGGT$_{Kif}$ phase determination: molecular replacement confirms $P2_12_12_1$, 1 copy/asu and 62% solvent

CCP4-Phaser was run in all primitive orthorhombic space groups searching with a copy of PDB ID 5NV4 from which TRXL2 was removed (declaring a RMSD of 2.0 Å - Phaser refined it to 0.77 Å). The results were clearly best in $P2_12_12_1$ (RF Z-score 7.0; TF Z-score: 10.4; Refined TFZ-equiv: 16.3; LLG: 114; Refined TF Z-score: 16.3, Refined LLG: 208. wR=0.58.5).

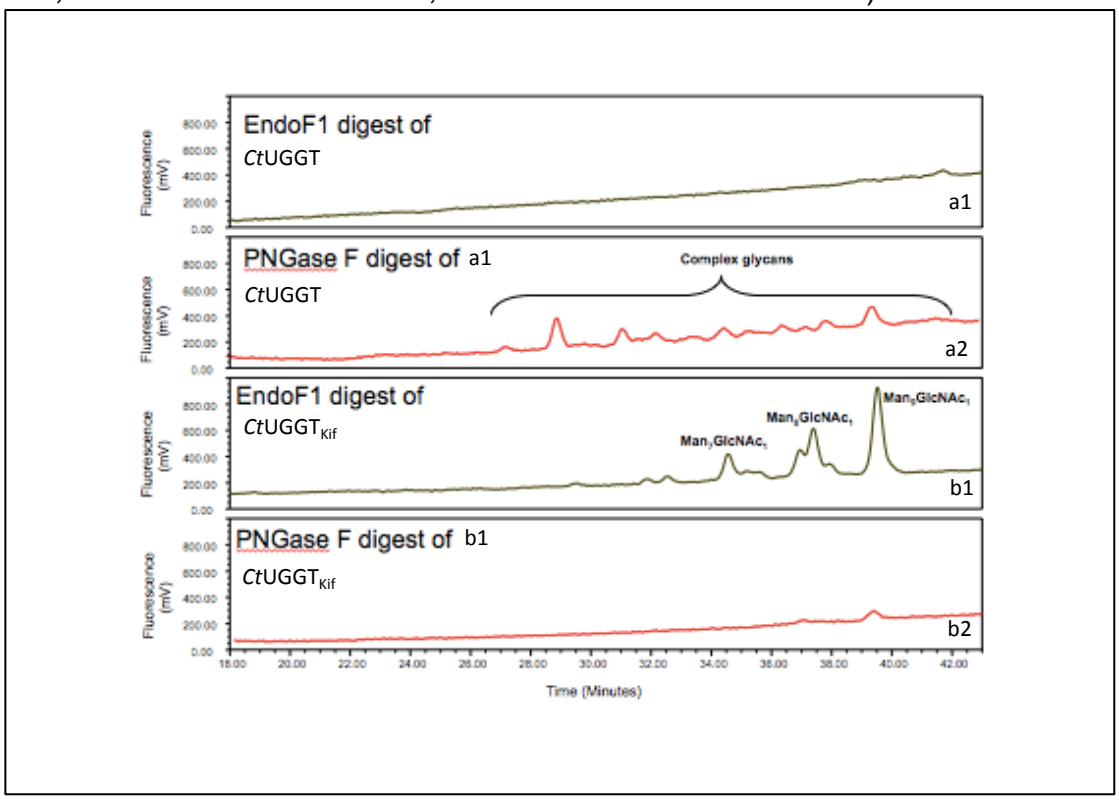


**Figure 3. *Ct*UGGT$_{\alpha}$ sequential digestions with Endo-F1 and PNGaseF.** After 2AA labelling, the glycans were analysed on an HPLC. **a1.** Glycans after Endo-F1 treatment of *Ct*UGGT: no peaks (all glycans are complex ones); **a2.** Glycans after PNGaseF treatment of the Endo-F1 treated *Ct*UGGT: complex glycans; **b1.** *Ct*UGGT$_{\alpha}$ carries N-linked glycans that EndoF1 is able to cleave, therefore they must be high-mannose glycans; **b2.** Glycans after PNGaseF treatment of the Endo-F1 treated *Ct*UGGT$_{\alpha}$: as Endo-F removed the high-mannose glycans, no glycans are left.

## *Ct*UGGT$_{Kif}$ refinement and model building

The first map obtained in autoBUSTER from this MR model (which lacks TRXL2) showed strong density for the TRXL2 domain. The TRXL2 domain was added by superposing PDB ID 5NV4 onto the model, and real-space fitting the domain to the Fo-Fc map in CCP4-coot.

The structure was refined in autoBUSTER (5) with one TLS body per domain, one rigid body per domain, with external restraints (6) to PDB ID 5NV4. As an example, the last autoBUSTER refinements were run with the following options:

```
refine -m Batches-2-3-4-5ShortLambda.P212121_sftools.nosysabs.mtz \
-p Ninth-Target-2cycles.pdb\ -target 5nv4.NoGlycans.pdb \
-TLS CtUGGT.tls \
-d Tenth-Deposition StopOnGellySanityCheckError=no \
AnalyseGellySanityCheckForDuplicateBonds=no \
-Seq CtUGGT.seq
```

**Table 1.** *CtUGGT$_{Kif}$* data processing statistics. In parentheses the values referring to the highest resolution shell.

| **Structure** | ***Ct*UGG$_{Kif}$ `New Intermediate' PDB ID 6TRF** |
|---|---|
| **Experiment** | I03@DLS 01.05.2016 |
| **Space group (Z)** | $P2_12_12_1(4)$ |
| **Wavelength (Å)** | 0.97630 |
| **Cell dimensions *a, b, c* (Å)** | a=78.65, b=148.93, c=190.30 |
| **α, β, γ (°)** | 90,90,90 |
| **Resolution range(Å)** | 95.15-4.11 (4.49-4.11) |
| **Measured reflections** | 82,040 (4,313) |
| **Unique reflections** | 7,500 (271) |
| $R_{merge}$ | 0.149 (3.268) |
| $R_{meas}$ | 0.157 (3.376) |
| $CC_{1/2}$ | 0.997 (0.566) |
| $I/\sigma I$ | 9.9 (1.3) |
| **Completeness (%)** | 90.2 (74.2) |
| **Redundancy** | 10.9 (15.9) |

Table 2 reports the Rfactors and geometry statistics at the end of the model building and refinement.

| Crystal form | *Ct*UGGT$_{Kif}$ |
|---|---|
| **PDB ID** | 6TRF |
| **Space group (Z)** | $P2_12_12_1$ (4) |
| **Resolution (Å)** | 95.1-4.1 (4.6-4.1) |
| **No. reflections** | 7,503 (442) |
| $R_{work}$ / $R_{free}$ | 0.25/0.31 (0.26/0.22) |
| **Atoms** | 10,717 |
| ***B*-factors (Å$^2$)** | 270 |
| **Rmsd$_{bonds}$(Å)** | 0.006 |
| **Rmsd$_{angles}$ (°)** | 0.95 |
| **Ramachandran outliers** | 17/1309 (1.3%) |
| **Ramachandran allowed** | 1292/1309 (98.7%) |
| **Ramachandran favoured** | 1235/1309 (94.3%) |

Figure 4 shows the *CtUGGT$_{Kif}$* structure (white) overlaid using Theseus(7) onto PDB ID 5NV4 (*aka Ct*UGGT-H, D611C:G1050C**,** rainbow). Rmsd$_{Cα}$ 2.2 Å over 1235 Cαs. *Ct*UGGT$_{Kif}$ adopts a conformation which combines a TRXL1-TRXL3 distance as found in the 'open' conformation (TRXL1 blue, TRXL2 green/cyan), but a TRXL2/TRXL3 relative orientation similar to the one found in the 'closed-like' conformation (TRXL3 yellow/green). We label this *Ct*UGGT$_{Kif}$ conformation 'new intermediate'.

Figure 5 shows the TRXL2 domains of *CtUGGT$_{Kif}$* and PDB ID 5NV4: it is apparent that the central sheet is absent in the former model. Residues *CtUGGT$_{Kif}$* 462-505 (two strands and one alpha helix) are not traceable in the electron density of the *CtUGGT$_{Kif}$* crystal. Whether this disorder/misfolding is correlated with the kifunensine treatment remains to be investigated. Indeed, the first steps of commitment of a terminally misfolded glycoprotein to ER associated degradation (ERAD) is digestion on the part of EDEM ER mannosidases of mannoses on the N-linked glycan: if the *CtUGGT* TRXL2 domain is prone to misfold, kifunensine-inhibition of ERAD may allow secretion of *CtUGGT$_{Kif}$* molecules that would normally be degraded by ERAD. By purifiying protein from kifunensine treated cells, we may have been able to capture a physiologically misfolded *Ct*UGGT (normally undergoing ERAD) in the crystals. NMR studies of UGGT TRXL2 domains are in progress, to check the fold of the TRXL2 domain in solution as a function of the temperature.

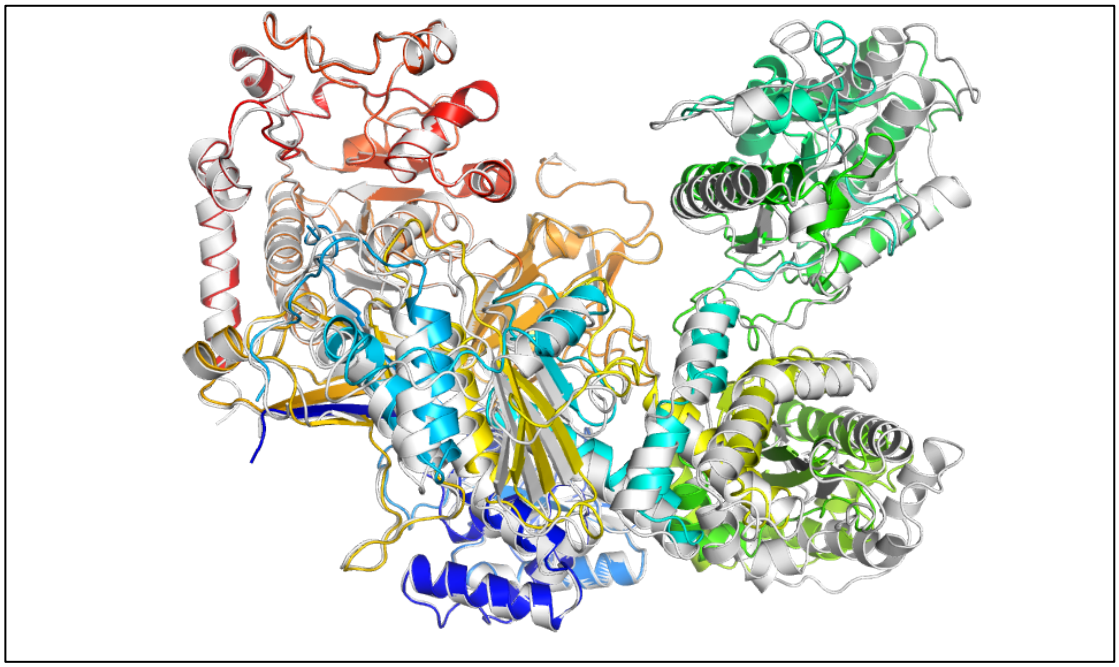

**Figure 4. *CtUGGT$_{Kif}$* structure.** *CtUGGT$_{Kif}$* structure (white) overlaid using Theseus(7) onto PDB ID 5NV4 (*aka Ct*UGGT-H, D611C:G1050C, rainbox, blue to red from N- to C-term).

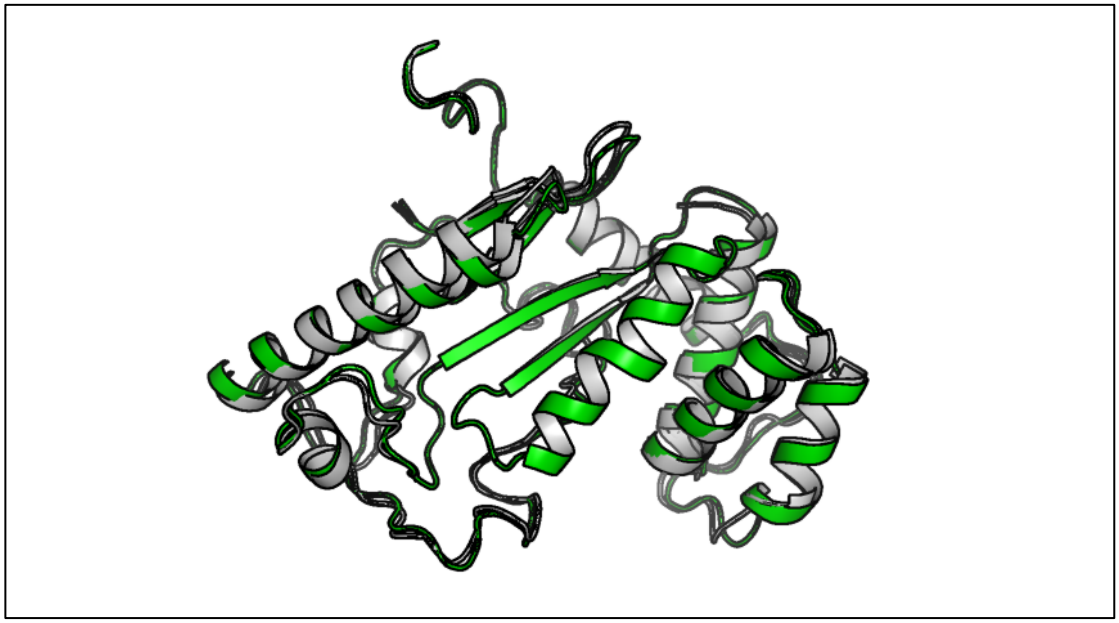

**Figure 5.** *Ct*UGGT$_{Kif}$ TRXL2 domain (from PDB ID 6TRF). *Ct*UGGT$_{Kif}$ TRXL2 (white) overlaid using Theseus(7) onto PDB ID 5NV4 TRXL2 (green). Residues *CtUGGT$_{Kif}$* 462-505 (two strands and one alpha helix, in the centre of the figure) are disordered in the crystal and missing from the model.

## References

1. Roversi P, et al. (2017) Interdomain conformational flexibility underpins the activity of UGGT, the eukaryotic glycoprotein secretion checkpoint. *Proc Natl Acad Sci USA*. doi:10.1073/pnas.1703682114.

2. Grueninger-Leitch F, D'Arcy A, D'Arcy B, Chène C (1996) Deglycosylation of proteins for crystallization using recombinant fusion protein glycosidases. *Protein Sci* 5(12):2617–2622.

3. Caputo AT, et al. (2018) Structural Insights into the Broad-Spectrum Antiviral Target Endoplasmic Reticulum Alpha-Glucosidase II. *Adv Exp Med Biol* 1062(8):265–276.

4. Vonrhein C, et al. (2011) Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr D Biol Crystallogr* 67(Pt 4):293–302.

5. Bricogne G, et al. (2017) BUSTER 2.10.3.

6. Smart OS, et al. (2012) Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta Crystallogr D Biol Crystallogr* 68(Pt 4):368–380.

7. Theobald DL, Steindel PA (2012) Optimal simultaneous superpositioning of multiple structures with missing data. *Bioinformatics* 28(15):1972–1979.