

## How does an inspector call: structures of UGGT, the eukaryotic glycoprotein quality control checkpoint.

**Related publication:** Roversi P, Marti L, Caputo AT, Alonzi DS, Hill JC, Dent KC, Kumar A, Levasseur MD, Lia A, Waksman T, Basu S, Soto Albrecht Y, Qian K, Mclvor JP, Lipp CB, Siliqi D, Vasiljevic S, Mohammed S, Lukacik P, Walsh MA, Santino A, Zitzmann N. Interdomain conformational flexibility underpins the activity of UGGT, the eukaryotic glycoprotein secretion checkpoint. *Proceedings of the National Academy of Sciences* **114**(32), 8544-8549, doi:10.1073/pnas.1703682114 (2017).

**Publication keywords:** UGGT; Endoplasmic reticulum; Glycoprotein folding; UDP-glucose:glycoprotein glucosyltransferase; Eukaryotic secretion

**I**n all cells derived from animals, plants, fungi and protozoa a single protein called UGGT acts as a quality control checkpoint that inspects secreted glycoproteins. UGGT has fascinated scientists for many years as it has the capacity to check the quality of a huge array of glycoproteins, which come in many different shapes and sizes and folds. The way in which it screens such a diverse group of proteins is poorly understood, and no structures exist of UGGT. Furthermore, UGGT is implicated in human diseases as it can ensure that glycoproteins from viruses (such as influenza, hepatitis C, HIV and Zika) are folded correctly. By fully characterising UGGT, it is hoped that an inhibitor could be developed that may act as a broad-spectrum anti-viral drug.

Using the Macromolecular Crystallography (MX) beamlines (I02, I03, I04 and I04-1) and cryo-electron microscopy (cryo-EM) at the Electron Bio-Imaging Centre (eBIC), the structure of UGGT was solved. Four distinct crystal structures of a full-length UGGT revealed a novel seven domain fold, with four thioredoxin-like domains arranged in a long arc, which terminates in two  $\beta$ -sandwich domains that tightly clasp the catalytic domain. Together, the cryo-EM map, the crystal structures and the biochemical data suggested that UGGT is highly flexible, a feature which enables it to check such a wide variety of glycoproteins.

A wonderfully efficient protein folding machinery in the Endoplasmic Reticulum (ER) of eukaryotic cells ensures that only correctly folded glycoproteins can exit the ER and proceed to the Golgi, and from the Golgi continue along the secretory pathway towards their individual cellular or extracellular destinations.

The stringency of ER Quality Control (ERQC) – while of great advantage to healthy cells – can lead to severe disease in patients carrying certain

mutant glycoproteins: the latter are detected as incorrectly folded, and either retained in the ER or degraded by ER-associated degradation (ERAD). The zeal of the otherwise excellent ERQC machinery bears particularly unfortunate consequences when the mutation affecting the glycoprotein is mild, i.e. it impairs but does not abrogate its function ('responsive mutant'). In these cases, ERQC blocks mutated competent glycoprotein secretion - and disease ensues.

Figure 1: The 15.2 Å cryo-EM CtUGGT reconstruction confirms mobility of TRXL2. Overlay of 3 crystallographically independent molecules of WT CtUGGT across 3 crystal forms. Helices are represented by cylinders. Yellow: 'open' conformation; green: 'intermediate' conformation; blue 'closed' conformation. The models have been superimposed with the main class of the 15.2 Å cryo-EM reconstruction. The cryo-EM reconstruction contour level encloses a volume corresponding to the protein mass.

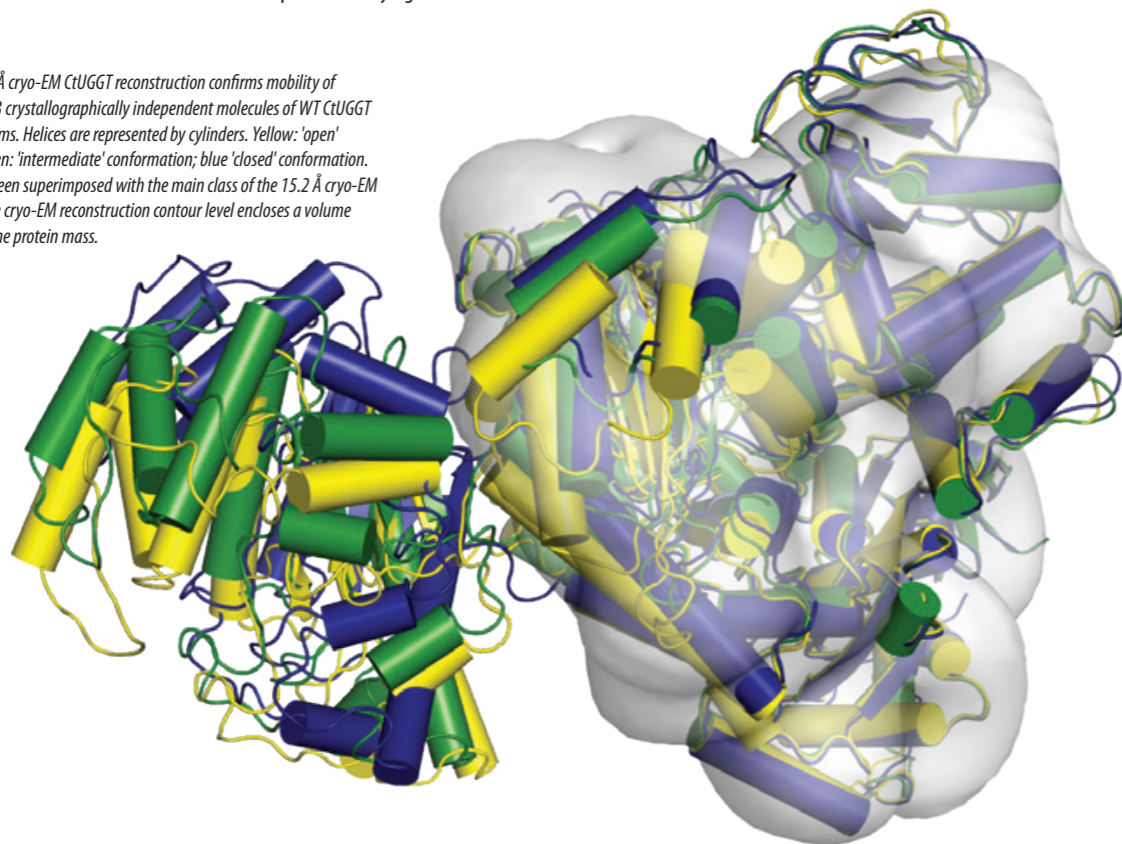


Figure 2: The 3D-printer generated CtUGGT model with its seven domains coloured TRXL1 (purple), TRXL2 (blue), TRXL3 (cyan), TRXL4 (green),  $\beta$ S1 (yellow),  $\beta$ S2 (orange) and GT (red). (a) a large, white LEGO glycoprotein; (b) a medium-size magenta LEGO glycoprotein; (c) a small, purple LEGO glycoprotein. Jelly babies at the end of a piece of string symbolise the Man9GlcNAc2 N-linked glycans on the glycoproteins. Disclaimer: the relative size of glycan to protein is not to scale.

In rare disease patients carrying responsive glycoprotein mutants, it is postulated that small molecule modulators of ERQC would rescue the secretion and residual activity of slightly misfolded and yet still active glycoproteins, thus alleviating pathogenic symptoms<sup>1</sup>. The fraction of responsive patients who could derive therapeutic benefit from ERQC modulation therapy depends on the disease, and varies between 70% in cystic fibrosis patients to 15-50% in lysosomal storage diseases<sup>2</sup>.

A second important field of medicine in which ERQC modulation may be of therapeutic value is virology, with many enveloped viruses (e.g. flu, Hepatitis C, HIV, and a number of tropical viruses such as Zika and dengue) needing the help of ERQC in order for their glycoproteins to fold correctly, and for infectious viruses to be successfully secreted from an infected cell.

At the heart of ERQC is the glycoprotein folding quality control checkpoint, the enzyme UDP-glucose:glycoprotein glucosyltransferase (aka UGGT). The enzyme selectively re-glucosylates misfolded glycoproteins, promotes their association with ER lectins and associated chaperones and prevents premature secretion from the ER<sup>3</sup>. UGGT has long resisted structural determination and sequence-based domain boundary predictions. It is currently still a big mystery how the many UGGT substrate glycoproteins, each of which has a different size, shape and fold, can all be surveyed by UGGT alone.

Before our study, no full-length UGGT structures were known<sup>4</sup>; nor to this date are UGGT inhibitors known (other than its product UDP). The protein sequence of UGGT does not resemble any other in the eukaryotic genome, so any structural details about the enzyme could shed light on UGGT function and inform ongoing efforts to develop UGGT inhibitors. We set out to characterise UGGT both structurally (X-ray crystallography, SAXS, cryo-EM) and biochemically.

Our crystal structures of a full-length UGGT revealed a novel seven domain fold, with four thioredoxin-like (TRXL) domains arranged in a long arc, which terminates in two  $\beta$ -sandwiches tightly clasping the glucosyltransferase domain. The presence of TRXL domains in the N-terminal part of UGGT brings to mind bacterial periplasmic disulphide isomerases and eukaryotic ER luminal chaperones binding misfolded proteins and/or peptides<sup>5</sup>.

Together with the 15 Å cryo-EM map, the same UGGT crystal structures reveal pronounced inter-domain flexibility (Fig. 1). Structure-based double cysteine point mutants that engineer extra inter-domain disulfide bridges, rigidify the UGGT structure and exhibit impaired activity. Thus, UGGT enzymatic activity depends on inter-domain conformational mobility, suggesting UGGT's intrinsic flexibility as a key requirement for the enzyme to be able to recognise and react with its many different misfolded substrates, flagging them for retention. By changing the relative orientations between its four TRXL domains, the enzyme could adapt its conformation to fit the particular glycoprotein it has bound and needs to re-glucosylate.

Structures of complexes of a UGGT with substrate misfolded glycoproteins

would help verify these ideas, and we have started work towards biochemical purification and structural elucidation of such UGGT:misfolded glycoprotein complexes. In the meantime, in order to illustrate and communicate the novel hypothesis that the enzyme's flexibility may be necessary to enable substrate promiscuity, we have generated a 3D -printed molecular representation of CtUGGT, and assembled three distinct LEGO generic glycoproteins models, each of different size and shape, and bearing one jelly-baby 'glycan'. Figures 2a-c show the 3D-printer UGGT model in three arbitrary conformations, with the TRXL domains (coloured purple, blue, cyan and green) grabbing hold of the LEGO glycoproteins. The latter are oriented in such a way that the jelly-baby "glycan" is presented to the UGGT catalytic glucosyltransferase domain, coloured in red.

### References:

- Amara, J. F. *et al.* Intracellular protein trafficking defects in human disease. *Trends in Cell Biology* **2**(5), 145-149, doi:10.1016/0962-8924(92)90101-R (1992).
- Parenti, G. *et al.* Pharmacological Chaperone Therapy: Preclinical Development, Clinical Translation, and Prospects for the Treatment of Lysosomal Storage Disorders. *Molecular Therapy* **23**(7), 1138-1148, doi: 10.1038/mt.2015.62 (2015).
- Parodi, A. J. *et al.* UDP-Glucose: Glycoprotein Glucosyltransferase 1,2 (UGGT1,2). *Handbook of Glycosyltransferases and Related Genes*, Springer 15-30, doi:10.1007/978-4-431-54240-7\_107 (2014).
- Zhu, T. *et al.* Structural insight into substrate recognition by the endoplasmic reticulum folding-sensor enzyme: crystal structure of third thioredoxin-like domain of UDP-glucose:glycoprotein glucosyltransferase. *Scientific Reports* **4**(7322) doi: 10.1038/srep07322 (2014).
- Kozlov, G. *et al.* A structural overview of the PDI family of proteins. *The FEBS Journal* **277**, 3924-3936, doi: 10.1111/j.1742-4658.2010.07793.x (2010).

### Funding acknowledgement:

This work was supported by the Oxford Glycobiology Endowment. JH and ATC were supported by Wellcome Trust Structural Biology Studentships (numbers ?). Italian funding?

**Corresponding authors:** Dr Pietro Roversi and Prof Nicole Zitzmann, Oxford Glycobiology Institute, University of Oxford, pietro.roversi@bioch.ox.ac.uk, nicole.zitzmann@bioch.ox.ac.uk.