Structural insight into substrate recognition by the endoplasmic reticulum folding-sensor enzyme: crystal structure of third thioredoxin-like domain of UDP-glucose:glycoprotein glucosyltransferase

Tong Zhu1,2,3, Tadashi Satoh3,4 & Koichi Kato1,2,3

1School of Physical Sciences, The Graduate University for Advanced Studies, 5-1 Higashiyama, Myodaiji, Okazaki, Aichi 444-8787, Japan, 2Okazaki Institute for Integrative Bioscience and Institute for Molecular Science, National Institutes of Natural Sciences, 5-1 Higashiyama, Myodaiji, Okazaki, Aichi 444-8787, Japan, 3Graduate School of Pharmaceutical Sciences, Nagoya City University, 3-1 Tanabedori, Mizuho-ku, Nagoya 467-8603, Japan, 4JST, PRESTO, 3-1 Tanabedori, Mizuho-ku, Nagoya 467-8603, Japan.

The endoplasmic reticulum (ER) possesses a protein quality control system that supports the efficient folding of newly synthesized glycoproteins. In this system, a series of N-linked glycan intermediates displayed on proteins serve as quality tags. The ER folding-sensor enzyme UDP-glucose:glycoprotein glucosyltransferase (UGGT) operates as the gatekeeper for ER quality control by specifically transferring monoglucose residues to incompletely folded glycoproteins, thereby allowing them to interact with lectin chaperone complexes to facilitate their folding. Despite its functional importance, no structural information is available for this key enzyme to date. To elucidate the folding-sensor mechanism in the ER, we performed a structural study of UGGT. Based on bioinformatics analyses, the folding-sensor region of UGGT was predicted to harbour three tandem thioredoxin (Trx)-like domains, which are often found in proteins involved in ER quality control. Furthermore, we determined the three-dimensional structure of the third Trx-like domain, which exhibits an extensive hydrophobic patch concealed by its flexible C-terminal helix. Our structural data suggest that this hydrophobic patch is involved in intermolecular interactions, thereby contributing to the folding-sensor mechanism of UGGT.
UGGT acts as the gatekeeper in this system because this enzyme is capable of sensing the folding states of glycoproteins as potential substrates. UGGT only transfers monoglucose residues to incompletely folded glycoproteins. UGGT is a large enzyme, comprising approximately 1500 amino acid residues, which has been putatively divided into two regions: an N-terminal folding-sensor region, which accounts for approximately 80% of the enzyme and is not homologous with any known structures, and a C-terminal catalytic domain, which accounts for the remaining 20% of the enzyme and belongs to the glycosyltransferase family. However, no further structural information is available on this key enzyme to date. Thus, the structural basis of the working mechanism of the CNX/CRT cycle remains unclear.

In this study, to elucidate the working mechanism of UGGT, we attempted to characterize the three-dimensional (3D) structure of its N-terminal folding-sensor region. We selected Chaetomium thermophilum as the source organism for the structural study of UGGT. Our bioinformatics analyses predicted that the folding-sensor region of UGGT is divided into two regions: an N-terminal folding-sensor region, which accounts for approximately 80% of the enzyme and is not homologous with any known structures, and a C-terminal catalytic domain, which accounts for the remaining 20% of the enzyme and belongs to the glycosyltransferase family. However, no further structural information is available on this key enzyme to date. Thus, the structural basis of the working mechanism of the CNX/CRT cycle remains unclear.

In this study, to elucidate the working mechanism of UGGT, we attempted to characterize the three-dimensional (3D) structure of its N-terminal folding-sensor region. We selected Chaetomium thermophilum, a thermophilic fungus, which survives at temperatures of up to 60°C, as the source organism for the structural study of UGGT. Our bioinformatics analyses predicted that the folding-sensor region of UGGT contains three tandem thioredoxin (Trx)-like domains. Moreover, we determined the 3D structure of a Trx domain of UGGT, thereby providing structural insights into the mechanism of substrate recognition of this folding-sensor enzyme.

Results

Bioinformatics identification of three tandem Trx-like domains in folding sensor region of UGGT. To investigate the structure of the N-terminal folding-sensor region of UGGT, we subjected its amino acid sequence to bioinformatics analysis using the programs PSIPRED and DISOPRED. The results indicate that the folding-sensor region of UGGT exhibits well-formed secondary structures: a mixed α/β region in the N-terminal part and a β-sheet-rich region around the C-terminus. The secondary structure distributions appeared highly conserved across species. A remarkably disordered segment was identified at the connection between the β- and C-terminal catalytic domains. This structural feature is consistent with previously reported results of limited proteolysis.

Next, we attempted to identify structural domain(s) within the N-terminal folding-sensor region using InterPro and Phyre2. Regarding the β-domain, no significantly homologous domains were identified. On the other hand, the folding-sensor region of UGGT was found to contain three tandem Trx-like domains: Trx1, Trx2, and Trx3. Moreover, we determined the 3D structure of a Trx domain of UGGT, thereby providing structural insights into the mechanism of substrate recognition of this folding-sensor enzyme.
Trx3, 23.3%; Trx2 versus Trx3, 16.2% in *C. thermophilum*), suggesting variability in their three-dimensional structures.

**Crystal structure of the third Trx-like domain of UGGT.** Based on the bioinformatic prediction that folding-sensor region of UGGT possesses three tandem Trx-like domains, we performed bacterial expression, purification and crystallization of a series of Trx domains. First, we expressed each of the three Trx domains. Although we were able to express the Trx3 domain as a soluble protein, the Trx1 and Trx2 domains formed inclusion bodies in *Escherichia coli*. Therefore, we made tandem constructs for these constructs, we successfully crystallized the Trx3 domain with the optimization of its N- and C-terminal sequences (residues 671–831), based on the identification of proteolytically stable fragments. Of these constructs, we successfully crystallized the Trx3 domain with the optimization of its N- and C-terminal sequences (residues 671–831), based on the identification of proteolytically stable fragments. However, despite extensive trials, we were unable to obtain crystals of the tandem constructs Trx1-Trx2, Trx2-Trx3 or Trx1-Trx2-Trx3.

We determined two forms of the crystal structure of Trx3 domain at 3.4 and 1.7 Å resolutions. The final model of Form 1, refined to a resolution of 3.4 Å, had an $R_{work}$ of 23.5% and $R_{free}$ of 29.2% (Table 1). The crystal belonged to space group $I23$. The crystal belonged to space group $C221$, with a detergent ANAPOE C12E8 belonged to space group $C221$, and differed up to 1.70-Å resolution. In the crystal structure, one molecule was contained per asymmetric unit. The final model of Form 2 had an $R_{work}$ of 20.1% and $R_{free}$ of 24.6% (Table 1).

As expected from the bioinformatics analysis, the crystal structure displayed a typical Trx-like fold, i.e. a five-stranded β-sheet with a β1–β3–β2–β4–β5 arrangement surrounded by six α-helices (Fig. 1b and 1c). In the crystal structure, a part of β5–α6 loop (residues 816–818) was disordered. The C-terminal α6-containing segment showed a higher crystallographic B-factor (87.7 Å²) than the average value (79.7 Å²; Table 1). Comparison of the structure of the Trx3 domain of UGGT with known protein structures using the DALI server revealed that the protein disulfide bond isomerase (DsbA/C) homologue, *Salmonella enterica* ScsC, was the most structurally similar protein ($Z$-score = 9.4; RMSD = 2.9 Å; identify = 18.5%; PDB code: 4GZX). As representative of the DsbA/C structure, the well-characterized crystal structure of *E. coli* DsbC (PDB code: 1EEJ) was also shown in Supplemental Figure 2. The overall fold of Trx3 domain of UGGT was essentially identical to that of ScsC except for their variable α helical segments between 3 and 4 ($α3$ and $α4$ in UGGT-Trx3 and $α3$–$α5$ in ScsC) (Supplemental Fig. S2b). DsbC also share very similar fold with the UGGT Trx3 domain except for the N-terminal $α1$ helix, which directly follows the dimerization domain in DsbC, and variable $α3$/$α4$ helices (Supplemental Fig. S2c). Compared with the crystal structure of the *E. coli* thioredoxin trxA² (PDB code: 2TRX; Supplemental Fig. S2d), which exhibits typical Trx fold, three contiguous helical insertions, $α3$, $α4$ and $α5$, were identified between $β3$ and $β4$, as observed in DsbC²⁷. Furthermore, an N-terminal segment containing $α1$ and $β1$ regions of the Trx3 domain of UGGT was significantly different from that of the *E. coli* trxA² in terms of topological arrangement. In the folds shared by the Trx3 domain of UGGT, SscS and DsbC, $α1$ precedes $β1$, which makes anti-parallel β-strands with $β3$ (Supplemental Fig. S2a–c). In contrast, $α1$ was inserted between $β1$ and $β2$, both of which were parallel with respect to $β3$ (Supplemental Fig. S2d). In addition, our homology modeling suggest that the Trx1 and Trx2 domains exhibit typical Trx-like folds similar to the Trx3 domain and its structural homologs, except for the N-terminal and variable $α$ helical segments between 3 and 4 and an insertion loop (residues 226–293) in Trx1 (Supplemental Fig. S3).

The C-terminal α6 helix, which is followed by a putatively flexible linker region in UGGT, was completely disordered in the crystal structure of Form 2, suggesting the instability of this helix (Fig. 2b,
Because of the absence of the \(\alpha_6\) helix, an extensive hydrophobic patch was exposed on the surface of the Trx3 domain (Fig. 2b, centre). The detergent ANAPoE C12E8 was accommodated on this exposed hydrophobic patch. The \(\alpha_6\) helix was stabilized mainly through its hydrophobic surface, containing Phe820, Phe825, Phe828 and Leu829, which made contact with the hydrophobic patch, including Leu703 (\(\beta_2\)), Leu717, Phe724 (\(\alpha_2\)), Val804, Leu806 (\(\beta_4\)), Leu811 (\(\beta_5\)) and Ile814 (\(\beta_5-\alpha_6\) loop) (Fig. 2a, right). Most of these hydrophobic residues were involved in the interaction with the detergent in Form 2. Thus, the C-terminal \(\alpha_6\) helix and detergent molecule occupy the common hydrophobic surface of the Trx3 domain. These hydrophobic residues are highly conserved among species (Fig. 1 and Supplemental Fig. S1).

**Discussion**

In this study, we proposed that the folding-sensor region of UGGT contains three tandem Trx-like domains and, solved the first 3D structure of a structural domain, i.e. the third Trx-like domain, of this functional region (Fig. 1 and Supplemental Fig. S1). Trx-like domains are common to members of the protein disulfide isomerase (PDI) family, which are responsible for assisting protein folding in the ER\(^2\). Most PDI family members are multidomain proteins containing both redox-active and -inactive Trx-like domains in different arrangement\(^29,30\). For example, PDI (PDIA1) as a representative member of PDI family possesses four tandem Trx-like domains (designated \(a\), \(b\), \(b'\) and \(a'\), of which \(a\) and \(a'\) domains have a CXXC catalytic motif, whereas \(b\) and \(b'\) domains do not\(^31,32\). None of the Trx-like domains of UGGT possess the CXXC catalytic motif, indicating that this enzyme is not directly involved in thiol/disulfide exchange reactions. In this context, the cis-Pro loop adjacent to the CXXC motif, a hallmark of redox-active Trx-fold proteins\(^39\) and involved in substrate recognition in DsbA\(^30\), is not present in the Trx3 domain of UGGT. Noncatalytic Trx-like domains are often involved in substrate recognition\(^31,35\), co-factor interaction\(^36\) and functional intradomain interactions\(^4\). UGGT forms a stable complex with Sep15, a 15-kDa selenocystein-containing oxidoreductase\(^37\) which possesses one redox-active Trx-like domain and enhances the glucosyltransferase activity of UGGT\(^38\). It is plausible that Sep15 serves as a structural extension of UGGT with a complementary function.

Growing evidence implies that UGGT exhibits glucosyltransferase activity only against incompletely folded glycoproteins, suggesting that the folding-sensor region has exposed the hydrophobic patch as a principal substrate-binding site\(^7,12-14\). The Trx3 domain possesses an extensive hydrophobic patch, which is covered by the flexible C-terminal helix and can participate in interactions with hydrophobic molecules (Fig. 2). The hydrophobic residues involved in these intramolecular and intermolecular interactions are conserved across species (Supplemental Fig. S1). Thus, our crystallographic study provides an atomic view of the potential substrate-binding site of UGGT. In addition, our homology modeling data suggested that Trx1 and Trx2 domains also exhibit larger hydrophobic patches located at the opposite site as compared with that of the Trx3 domain, suggesting the possibility of their involvement in substrate recognition (Supplemental Fig. S4). Concomitantly, this may be the cause of inclusion body formation of the isolated Trx1 and Trx2 domains. In general, molecular chaperones undergo conformational transitions coupled with the shielding and exposure of their hydrophobic patches as substrate-binding sites\(^35,39\). Although we cannot exclude the possibility that the hydrophobic patch of the Trx3 domain is covered by other domain(s) in intact UGGT, the flexible properties of the C-terminal helix would be important for allowing the flexible properties of the C-terminal helix to interact with hydrophobic patches of substrate and other domains that could interact with these patches. Additionally, the flexible properties of the C-terminal helix could also allow for the flexible movement of the C-terminal helix, which may help to prevent the hydrophobic patch from being covered by other domains.
terminal helix of Trx3 may contribute to regulatory mechanisms underlying the folding-sensing function of this domain.

In summary, our bioinformatic analyses predicted that the folding-sensor region of UGGT harbours three tandem Trx-like domains. Moreover, we provided snapshots of the 3D structure of the third Trx-like domain, in which a putative substrate-binding hydrophobic patch is intramolecularly masked or involved in an intermolecular interaction, offering a key breakthrough toward understanding of the functional mechanisms of this ER folding-sensor enzyme.

**Methods**

**Protein expression and purification.** *C. thermophila var. thermophila* La Touche (DSM 1495) was obtained from DSMZ, Braunschweig, Germany. Total RNA was isolated using TRIzol® reagent (Life Technologies). The cDNA was synthesized using SuperScript® III Reverse Transcriptase (Life Technologies) with oligo d(T) primers according to the manufacturer’s instructions. Full-length UGGT cDNA was cloned by PCR using a *C. thermophila* genomic DNA database. Recombinant UGGT proteins were expressed as glutathione S-transferase (GST)-fused proteins. The Trx1 (residues 168–379), Trx2 (residues 467–624), and Trx3 (residues 671–831), Trx1-Trx2 (residues 168–624), Trx2-Trx3 (residues 467–831) and Trx1-Trx2-Trx3 (residues 168–831) domains were amplified by PCR and subcloned into the *BamHI* and *XbaI* sites of a modified pCold-GST vector (Takara Bio Inc.), in which the factor Xa site was replaced with the tobacco etch virus (TEV) protease recognition site. Recombinant proteins were expressed in *E. coli* BL21 Star™ cells (Life Technologies) according to the manufacturer’s protocols (Takara Bio Inc.). GST-fused proteins were purified using glutathione-Sepharose™ columns (GE Healthcare). Subsequently, the GST tag was removed by adding TEV protease to the resin for 12 h at 277 K, leaving two additional residues Gly-Ser at the N-terminals. The resultant proteins were further purified by size-exclusion chromatography (Superdex-200; GE Healthcare) using a buffer containing 20 mM Tris-HCl (pH 7.5), 150 mM NaCl and 0.1 mM EDTA. The seleneomethionine (SeMet)-labelled Trx3 domain was expressed in *E. coli* B834 (DE3) using M9 minimal medium with SeMet. Expression and purification were performed following the same protocol as that for the native protein. Purified proteins were dialyzed against a buffer containing 10 mM Tris-HCl (pH 7.5) and 100 mM NaCl. The integrity of the protein samples was validated by matrix-assisted laser desorption/ionization-time of-flight mass spectrometry (MALDI-TOF/MS) analysis using an AXIMA-CFR™ spectrometer (Shimazu) and N-terminal Edman sequencing with a Procise 494HT protein sequencer (AB/Life Technologies).

**Protein crystallization, X-ray data collection and structure determination.** The crystals of the Trx3 domain of UGGT (Form 1, 10 mg/ml) were grown in a buffer containing 60% Tacsimate (pH 7.0) for 2 weeks at 289 K. The crystals of the Trx3 domain of UGGT (Form 2) were obtained by equilibrating a solution of 8 mg/ml protein in a 2.2 mM ANAOPE C12E8 (polyoxyethylene[8]dodecyl ether) solution with a reservoir solution containing 3.6,9,12,15,18,21-hexaoxaclactone (1:1 volume) and 1% glicerol with an equal volume of precipitant solution containing 23% PEG3350, 0.1 M Tris-HCl (pH 7.0) and 0.2 M ammonium acetate for 6 days at 289 K. The crystals were transferred into the reservoir solution and flash-cooled in liquid nitrogen. Data sets for Forms 1 and 2 were collected using synchrotron radiation at 13B1 of the National Synchrotron Radiation Research Center (Hsinchu, Taiwan) and AR-NW12A of the Photon Factory (Tsukuba, Japan), respectively. All diffraction data were processed using HKL2000®. Crystal parameters are summarized in Table 1.

The 1.70 Å resolution crystal structure of the Trx3 domain of UGGT (Form 2) was solved using the SAD method. The initial phase was determined using the SHELX C/D/E program. Further manual model building into the electron density maps and refinement were performed using COOT® and REFMAC5®, respectively. The 3.40 Å resolution structure of the Trx3 domain of UGGT (Form 1) was solved by molecular replacement using the phase information from the lowest resolution structure. The crystallographic refinement refinement statistics are summarized in Table 1. Graphic figures were prepared using PyMOL (http://www.pymol.org/). Homology modeling of the Trx1 and Trx2 domains were performed using Phyre2 with *Neisseria gonorrhoeae* DsbC-like protein (PDB code: 3GV1) and *Neisseria meningitidis* meningitidis DsbA1 (PDB code: 3DVW) as templates, respectively.

1. Ellgaard, L. & Helenius, A. Quality control in the endoplasmic reticulum. *Nat Rev Mol Cell Biol* 6, 411–422 (2005).
2. Kato, K. & Kamiya, Y. Structural views of glycoprotein-fate determination in cells. *Glycobiology* 17, 1031–44 (2007).
3. Takeda, Y., Totani, K., Matsuo, I. & Ito, Y. Chemical approaches toward understanding glycan-mediated protein quality control. *Curr Opin Chem Biol* 21, 3152–62 (2010).
4. Erlandsson, K. G. Z. Glycoprotein folding, quality control and ER-associated degradation. *Curr Opin Struct Biol* 19, 515–23 (2009).
5. Aebi, M., Bernasconi, R., Clerc, S. & Molinari, M. N-glycan structures: recognition and processing in the ER. *Trends Biochem Sci* 35, 74–82 (2010).
37. Korotkov, K. V., Kumaraswamy, E., Zhou, Y., Hatfield, D. L. & Gladyshev, V. N. Association between the 15-kDa selenoprotein and UDP-glucoseglycoprotein glucosyltransferase in the endoplasmic reticulum of mammalian cells. J Biol Chem 276, 15330–6 (2001).
38. Takeda, Y. et al. Both isoforms of human UDP-glucoseglycoprotein glucosyltransferase are enzymatically active. Glycobiology 24, 344–50 (2014).
39. Saibil, H. Chaperone machines for protein folding, unfolding and disaggregation. Nat Rev Mol Cell Biol 14, 630–42 (2013).
40. Hayashi, K. & Kojima, C. pCold-GST vector: a novel cold-shock vector containing GST tag for soluble protein production. Protein Expr Purif 62, 120–7 (2008).
41. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. Methods in Enzymology 276, 307–326 (1997).
42. Sheldrick, G. M. A short history of SHELX. Acta Crystallogr A 64, 112–22 (2008).
43. Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. Nat Protoc 3, 1171–9 (2008).
44. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. Acta Crystallogr D Biol Crystallogr 66, 486–501 (2010).
45. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr 53, 240–55 (1997).
46. McCoy, A. J. et al. Phaser crystallographic software. J. Appl. Cryst. 40, 658–674 (2007).
47. Lovell, S. C. et al. Structure validation by Cα geometry: qα, qψ and Cβ deviation. Proteins 50, 437–50 (2003).
48. Pei, J., Kim, B. H. & Grishin, N. V. PROMALS3D: a tool for multiple protein sequence and structure alignments. Nucleic Acids Res 36, 2295–300 (2008).

Acknowledgments

We acknowledge Dr Hirokazu Yagi (Nagoya City University, Nagoya, Japan) and Ms Yumiko Makino (National Institute for Basic Biology, National Institutes of Natural Sciences, Okazaki, Japan) for help with MALDI-TOF/MS and N-terminal sequence analyses, respectively. This work was supported in part by the Okazaki ORION project and Grants-in-Aid for Scientific Research (Grant Numbers 24770102, 25121730 to T.S. and 25102008, 24249002 to K.K.) from the Ministry of Education, Culture, Sports, Science and Technology, Japan, and by PRESTO project (Grant Number 13417569 to T.S.) from the Japan Science and Technology Agency.

Author contributions

T.S. and K.K. conceived and designed the experiments; T.Z. and T.S. performed the bioinformatics analyses and crystallographic experiments; all authors wrote and reviewed the manuscript.

Additional information

Accession codes The coordinates and structural factors of the crystal structures of the Trx3 domain of C. thermophilum UGGT (Forms 1 and 2) have been deposited in the Protein Data Bank under the accession numbers 3WZT and 3WZS, respectively.

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Zhu, T., Satoh, T. & Kato, K. Structural insight into substrate recognition by the endoplasmic reticulum folding-sensor enzyme: crystal structure of third thioredoxin-like domain of UDP-glucoseglycoprotein glucosyltransferase. Sci. Rep. 4, 7322; DOI:10.1038/srep07322 (2014).