Functional analysis of endoplasmic reticulum glucosyltransferase (UGGT): synthetic chemistry's initiative in glycobiology

Yukishige Ito,^{1,2} Yoichi Takeda², Akira Seko,² Masayuki Izumi,³ and Yasuhiro Kajihara^{2,3}

¹Synthetic Cellular Chemistry Laboratory, RIKEN, Wako, Saitama 351-0198, Japan

²JST, ERATO, Ito Glycotrilogy Project, Wako, Saitama 351-0198, Japan

³Department of Chemistry, Graduate School of Science, Osaka University, Toyonaka, Osaka 560-0043, Japan

Fax: 81-48-462-4680, E-mail: yukito@riken.jp (Y. I.)

Abstract

UGGT1 is called as a folding sensor protein that recognizes misfolded glycoproteins and selectively glucosylates high-mannose-type glycans on the proteins. However, conventional approaches using naturally occurring glycoproteins is not optimum in performing precise analysis of the unique properties of UGGT1. We have demonstrated that high-mannose-type glycans, in which various hydrophobic aglycons were introduced, act as good substrates for UGGT1 and are useful analytical tools for its characterization. Moreover, we found that UGGT2, an isoform UGGT1, is also capable of glucosylating these synthetic substrates. Our strategy stemmed on synthetic chemistry has been further strengthened by total synthesis of homogeneous glycoproteins in correctly folded as well as in intentionally misfolded forms.

Keywords: UDP-Glc: glycoprotein glucosyltransferase; Glycoprotein quality control system; Synthetic oligosaccharides; Synthetic glycoproteins; Sep15.

Contents

1. UGGT in glycoprotein quality control system in the ER.....	
2. Synthesis of high-mannose-type glycans.....	
2.1 Convergent synthetic approach.....	
2.2 Top-down chemo-enzymatic approach.....	
3. Functional analysis of UGGT1 by in vitro assay.....	
3.1 Analyses using naturally occurring glycoproteins.....	
3.2 Analyses using synthetic substrates.....	
4. Substrate recognition of UGGT1.....	
4.1 The recognition of fluorophoric aglycon of synthetic substrates by UGGT1.....	
4.2 Chemical synthesis of glycoproteins as substrates of UGGT1.....	
4.3 Glucosylation of the glycoprotein folding intermediates.....	
5. Enzymatic activity of UGGT isoforms.....	
6. Perspective.....	
Acknowledgment.....	
References.....	

1. UGGT in glycoprotein quality control system in the ER

Glycoproteins produced in cells are highly diverse. To investigate their functions, isolation from natural sources has been an orthodox strategy. In most cases, however, they consist of glycoforms which differ in number, composition, branching, and terminal

modification of glycan chains. In consequence, isolation of a homogeneous glycoprotein with a defined structure is extremely difficult, unless the glycoprotein is exceptionally abundant. In order to advance the study of the biological roles of glycan chains, chemical synthesis is expected to be highly powerful, because it can provide homogeneous and structurally defined substrates in substantial quantities. It also enables the preparation of non-natural derivatives as required. These molecular probes are highly useful in gaining a precise understanding of the biological activities of oligosaccharides at the molecular level. This view is further strengthened by remarkable progress in glycoprotein synthesis, which has put cell-free production of biologically active glycoproteins into reality. In this article, we wish to summarize our recent study on a glucosyltransferase in the ER, which is one of key players in glycoprotein quality control system. We believe it showcases strength of research strategies stemmed on chemical synthesis in glycobiology.

In eukaryotic cells, the endoplasmic reticulum (ER) plays a central role as biosynthetic machinery that produces proteins by translation of mRNA. It is also a central arena of protein modifications, especially asparagine (Asn)-linked glycosylation (*N*-glycosylation) [1]. As a matter of fact, a majority of newly generated polypeptides delivered into the ER lumen through translocons are modified by *N*-glycosylation.

N-Glycosylation is catalyzed by a multisubunit membrane bound enzyme, oligosaccharyl transferase (OST) [2, 3]. The glycosylation usually takes place in a co-translational manner [4, 5], whereas occurrence of post-translational *N*-glycosylation [6] has been clarified by recent studies. In either case, the OST selects specific asparagine (Asn) residues that are embedded in the consensus sequence Asn-X-Ser/Thr ($X \neq \text{Pro}$) [7].

In most eukaryotes, OST transfers a oligosaccharide composed of 14 sugar residues consisting of 3 glucose (Glc), 9 mannose (Man), and 2 *N*-acetyl-D-glucosamine (GlcNAc) residues (Glc₃Man₉GlcNAc₂; G3M9), the largest among all “high-mannose-type” glycans in the ER. Curiously, thus introduced tetradecasaccharide is trimmed into smaller glycans by various glycosidases in the ER. This seemingly uneconomical process is now believed to be highly meaningful, because each subgroup of glycans interacts with distinct lectins or enzymes with high specificity (Figure 1A).

On this avenue, the G3M9 is first trimmed by glucosidase I (G-I), which removes the terminal Glc residue, generating Glc₂Man₉GlcNAc₂ (G2M9). Further trimming is mediated by glucosidase II (G-II), a dual activity enzyme, which is able to cleave Glc-Glc (cleavage 1) and Glc-Man (cleavage 2) from G2M9, leading the latter to monoglucosylated (Glc₁Man₉GlcNAc₂; G2M9) and eventually to non-glucosylated (Man₉GlcNAc₂; M9) glycans. All of these glycans are also subjects of trimming by various mannosidases which generates their congeners lacking one or more Man residues.

Several lines of evidence have suggested that the cleavage 1 of G-II proceeds much faster than the cleavage 2. Accordingly, it seems reasonable to speculate that glycoproteins having G1M9 are given longer life span than other glycoforms in the ER. Glycoproteins with this glycoforms enter the glycoprotein specific folding cycle, so-called calnexin (CNX)/calreticulin (CRT) cycle [8]. CNX and CRT are ER-residing chaperones characterized by their lectin-like property. They are known to specifically recognize glycoproteins that contain monoglucosylated high-mannose-type glycans, most typically G1M9.

This cycle is maintained by a pair of counteracting enzymes, G-II and UDP-Glc: glycoprotein glucosyltransferase 1 (UGGT1) [9]. G-II regulates both entry and exit of glycoproteins, due to its dual activity to generate and digest monoglucosylated (G1M9) glycoforms. The role of UGGT1 in the CNX/CRT cycle is highly intriguing. By virtue of its ability to discriminate folding states of glycoproteins, it is considered to play a role as the “folding sensor” in the ER. More specifically, UGGT1 “reglucosylates” folding defective glycoproteins that have been deglucosylated by G-II, while it is inactive toward folded glycoproteins, irrespective to their glycoforms. Thus, glycoproteins in the ER oscillate between mono- and non-glucosylate glycoforms as far as their folding has been incomplete, while correctly folding glycoproteins are ignored by UGGT1 and released from the cycle.

UGGT1 is an ER-residing soluble enzyme, which partially distributes in the ER-Golgi intermediate compartment (ERGIC). Wide distribution in various eukaryotes, such as animals, plants, fungi, and protozoans has been revealed [10, 11]. It is a relatively large protein, whose molecular weight is 150-170 kDa, having a signal peptide sequence in the N-terminus. Whereas an ER retention/retrieval signal is present in the C-terminus, the sequence differ slightly among species; REEL (Human), HEEL (Rat), HGEL (*Drosophila melanogaster*), HTEL (*Caenorhabditis elegans*), PDEL (*Schizosaccharomyces pombe*), and KAEL (*Arabidopsis thaliana*). The C-terminal region of UGGT1 is considered to be the catalytic region, which share similarity with glycosyltransferases belonging to family 8 (Figure 1B).

Despite its important role in the glycoprotein folding, precise analysis of UGGT1 has been difficult. Most of previous studies relied on the use of naturally occurring glycoproteins that have been artificially denatured. They include bovine thyroglobulin (Tg), soybean agglutinin (SBA), and bovine pancreatic ribonuclease B (RNase B) [12], among which Tg has been most often used. It contains 29 disulfide linkages and 13 *N*-linked glycans, 4 of which are high-mannose-type [13]. On the other hand, SBA is monoglycosylated, consisting mainly of a M9 glycan [14], which has no disulfide and exists as a tetramer. RNase B is a small glycoprotein with 124 AA, consisting of one high-mannose-type glycan. It is known to be a mixture of several glycoforms, among which M5 and M6 are most abundant [15], while the proportion of M9, which is known to be most reactive as a UGGT1 substrate, is low.

In all cases, these glycoproteins are denatured chemically in order to be used as substrates of UGGT1. For instance, treatment with 8 M urea at room temperature has been employed to denature Tg, while more elevated temperature is required for SBA.

In spite of significant effort, mechanistic understanding of its unique property has been poorly made [9]. Intriguingly, UGGT1 seems to discriminate subtle conformational differences of proteins. While it does not accept folded glycoproteins, extensively misfolded glycoproteins are also poor substrates of this enzyme [16]. Instead, UGGT1 prefers proteins with molten-globule like structure as substrates [17, 18]. As such, the question as to how the UGGT1 discriminate folding state of glycoprotein has remained unanswered. [19]

Several studies have revealed the dual recognition mode of UGGT1. In addition to folding defective region of proteins, the innermost GlcNAc is strictly required in order for the substrate glycoprotein to be recognized by UGGT1 [12, 20, 21]. This notion was first formulated by a study of Ritter *et al.*, who employed modified ribonuclease (RNase) as substrates or inhibitors. They discovered that misfolded RNaseB that retained GlcNAc exhibited inhibitory activity toward UGGT1, while misfolded RNaseA was inactive, indicating that the presence of innermost GlcNAc is indispensable for binding [22].

2. Synthesis of high-mannose-type glycans

2.1 Convergent synthetic approach

Analysis of molecular interactions in the glycoprotein quality system has been difficult, mainly because of limited availability of substrates. As naturally occurring glycoproteins are usually heterogeneously glycosylated, they are not optimum as substrates. Although glycans released from proteins can be labeled and purified into high homogeneity, their variation is still limited. In particular, as UGGT1 requires both misfolded polypeptide and Asn-linked GlcNAc, commercially available PA sugars are of limited use.

In view of that, our study first aimed to develop systematic synthetic routes to high-mannose-type glycans. Although fully chemical synthesis of biologically relevant glycans consisting of more than ten sugar residues is not a trivial task, comprehensive synthesis of high-mannose-type glycans was achieved based on a convergent strategy (Figure 2A). Namely, combination of fragments corresponding to core trisaccharide (ManGlcNAc₂; **A**), trimannose (Man₃; **B**), branched mannose (Man₅₋₃; **C1,2,3,4**) and

glucose (Glc₃₋₁; **D1,2,3**) gave all possible high-mannose-type glycans [23, 24]. In addition, a uniform protocol was established for further modification, which consists of 1) conversion to glycosylamine, 2) introduction of a glycine linker, and 3) introduction of substituents that enable spectroscopic detection and protein conjugation.

In our study, glycans modified by methotrexate (MTX) were shown to be highly versatile. Namely, MTX conjugated glycans (CHO-MTX) were easily detectable because of intense UV absorbance of MTX [25]. Furthermore, as MTX is a tight binding inhibitor of dihydrofolate reductase (DHFR), DHFR grafted by various glycans (CHO-MTX-DHFR) can be prepared simply by mixing CHO-MTX and DHFR (Figure 2B) [26]. Alternatively, introduction of fluorescently active substituents was achieved in a similar manner.

2.2 Top-down chemo-enzymatic approach

Although our convergent strategy is versatile, one by one syntheses of diverse glycans is inevitably time consuming and labor intensive. To alleviate this problem, our recent study explored to develop top-down chemoenzymatic approaches that utilize enzymatic trimming of a single precursor (Figure 3). For this purpose, we designed a non-natural tetradecasaccharide **I** as the common precursor, which was chemically synthesized in a convergent manner [27]. A design was made to cap each branch of the glycan by a distinct sugar residue, which can be removed without risking rest of the molecule. Specifically, we employed Glc, Gal, and GlcNAc residues to cap A-, B-, and C-arms, respectively. It was converted to as many as 35 glycans by treatment with logical sets of glycosidases, greatly speeding up the preparation of high-mannose-type glycans and their non-natural congeners.

Conversion of **I** to all type of mono- and non-glucosylated high-mannose type glycans was achieved by using logically designed set of enzymes. For instance, preparation of Man8C from **I** was conducted by treatment with α 1,2-mannosidase after removal of GlcNAc on C-arm, which preserved α 1,2-mannosyl linkages on the A- and B-arms. Moreover, stepwise trimming of two sequential α 1,2-mannosyl linkages on the A-arm is achievable using Golgi endo- α -mannosidase, which cleaves a disaccharide (Glc α 1,3Man) from monoglucosylated oligosaccharides, and α 1,2-mannosidase.

3. Functional analysis of UGGT1 by in vitro assay

3.1 Analyses using naturally occurring glycoproteins

As a folding sensor in the ER, UGGT1 discriminates conformational status of newly generated proteins. Namely, it reglucosylates glycoproteins that have lost all Glc residues by G-II digestion, only when the folding is incomplete. Whereas the role of UGGT1 in CNX/CRT cycle seems to be perfectly purposive in glycoprotein quality control, mechanistic understanding of its unique property has been difficult to attain. Nevertheless, quite extensive studies based on in vitro experiments have elegantly revealed key features of this enzyme. For instance, the study by Trombetta et. al., which employed Tg revealed that Tg was a substrate of UGGT1 when it was exposed to denaturing conditions, such as 8 M urea, while native Tg was inactive [28]. On the other hand, Sousa et. al. revealed that UGGT1 preferentially recognizes GlcNAc residue of N-glycans covalently attached to proteins having non-native conformation [20]. Similarly, a study by Ritter *et al.*, which employed variously modified ribonuclease A and B, suggested that oligosaccharide fine

structure is not of primary importance, whilst the presence of innermost GlcNAc is indispensable for binding [22].

By using SBA as a model substrate, Caramelo et al. explored how the enzyme recognizes oligomeric state of proteins. The native SBA, which exists as a tetramer, dissociates into monomer upon treatment with 2 M urea. Although the monomer retained its native conformation, it was shown to be active as a substrate of UGGT1. They suggested that SBA exposed its hydrophobic interface by dissociation [29]. On the other hand, Zapun et al. examined several conformers of RNase B as substrates. In this case, unfolded RNase B (U-RNase B) obtained by denaturation was reduced and alkylated, either by *N*-ethylmaleimide or iodoacetamide. A mixture of folding intermediates (I-RNase B) was also generated artificially by oxidative refolding and alkylation of reduced RNase B [30].

Trombetta and Helenius developed RNase BS, by subtilisin treatment of RNase B, which introduced a nick between Ala20 and Ser21 of the protein. Whereas RNase BS keeps some conformational elements of the native protein, it was shown to be a better substrate of UGGT1 than fully unfolded RNase B. This observation indicates that UGGT1 is able to detect quite subtle conformational deviation [21].

3.2 Analyses using synthetic substrates

In spite of these important discoveries, precise understanding as to how UGGT1 discriminates folding states is still enigmatic. In addition, as most or previous studies employed naturally occurring glycoproteins as substrates, relative reactivity of glycoforms has been unclear. We adventitiously found that MTX-modified M9 (M9-MTX) was an excellent substrate of UGGT1 [25]. It may be speculated that M9-MTX mimics certain

property of misfolded glycoproteins, most likely hydrophobicity. As a matter of fact, unsubstituted M9 was barely reactive, while *n*-propyl glycoside of M9 exhibited reactivity in between M9-MTX and M9. Interestingly, conjugation with DHFR abrogated reactivity of M9-MTX, indicating that M9-MTX-DHFR behaved similarly to folded glycoproteins (Figure 4A).

Subsequent preparation of various glycans modified by MTX gave us quantitative estimate of glycan specificity of UGGT1. The order of reactivity was shown to be M9>M8>M7, relative values of initial velocity being 100, 56, and 20, respectively. Interestingly, although its reactivity was low, M7 exhibited highest affinity to UGGT1 (smallest K_m). Interestingly, a glycans lacking the innermost GlcNAc was totally devoid of activity, proving that the innermost GlcNAc is indispensable to be recognized by UGGT1 (Figure 4B) [25].

Further study found a variety of non-proteinic substrate of UGGT1. For instance, glycans modified by fluorophoric substituents, such as BODIPY or TAMRA were shown to have reactivity exceedingly higher than M9-MTX. Since then, these compounds have been used routinely for high-sensitivity detection of the UGGT1 activity [31].

For example, we have examined if UGGT1 is permissive to modification of donor substrate. It was shown to incorporate galactose, glucuronic acid, and 6-deoxy-glucose from corresponding UDP-sugars, although the conversion efficiency was low [32]. Interestingly, UDP-5-thio-glucose was revealed to be a far more reactive, whose V_{max}/K_m value being ca. 0.25 of that of UDP-Glc [Sakono, Takeda, et al., unpublished].

Although why some non-proteinic glycans exhibited high-reactivity to UGGT1 is hard to pin down, the easiest interpretation would be that these substrates mimic key element of unfolded proteins, such as surface exposed hydrophobicity [17, 18]. Whatever the mechanistic rationalization is, the advantage of using synthetic substrates for analysis of UGGT1 is obvious, because of their glycan structural homogeneity.

4. Substrate recognition of UGGT1

4.1 The recognition of fluorophoric aglycon of synthetic substrates by UGGT1

With versatile synthetic substrates of UGGT1 in hands, we turned our focus on the folding sensing mechanism of this enzyme. To this aim, fluorescently labeled compounds were particularly valuable. In order to evaluate the effect of distance between glycan and hydrophobic aglycon, we prepared a series of M9 derivatives linked to BODIPY by PEG spacers of various lengths (Figure 5A) [31]. The analysis revealed that the reactivity gradually decreased as the length of the spacer increased. However, interpretation of this result requires caution, as hydrophobicity of the aglycon might have been compromised more significantly as length of the PEG spacer increases. In our view, the results rather seem to suggest conformational flexibility of aglycon recognizing domain of the enzyme, which is capable to accept various glycoproteins with variable distance and/or orientation between their glycans and misfolded region.

The role of hydrophobic interaction in substrate recognition was reinforced by an ANS binding experiment. ANS is a fluorescent probe whose intensity increases when it associates with hydrophobic region of protein surface. In the presence of UGGT1, the fluorescent intensity increased and the maximum emission wavelength shifted from 515 nm

to 488 nm [33]. In addition, incubation with M9-G-E-TAMRA, an excellent substrate of UGGT1, caused significant change of circular dichroism (CD) spectrum of UGGT1 (Figure 5B). Either M9-glycan or 5-TAMRA exhibited no effect, suggesting that the conformational change was a result of the enzymes ability to simultaneously recognize both the glycan and the hydrophobic aglycon.

4.2 Chemical synthesis of glycoproteins as substrates of UGGT1

In order to analyze events related to glycoprotein quality control, naturally occurring glycoproteins have been conventionally employed. However, they often have limitations associated with their heterogeneity in terms of glycan occupancy and structure and conformational status of proteins. On the other hand, our study has been conducted mainly with synthetic glycans artificially modified by small molecules as aglycons. In this case, however, concern always arises as for biological relevance of the system. With these problems in mind, the use of chemically synthesized homogeneous glycoproteins would be the best solution. Although total synthesis of glycoproteins was considered to be desperately challenging, recent progress in this field has been astonishing. In fact a number of biologically important glycoproteins have been synthesized chemically [34].

As substrates of UGGT1, we explored synthesis of proteins modified by high-mannose-type glycans [35]. In this case, they selected interleukin 8 (IL-8) as a model, which was modified in its Asn36 by a M9 glycan (M9-IL-8). The protein consists of 72 amino acid residues with two disulfides between Cys7-Cys34 and Cys9-Cys50. To achieve the synthesis, native chemical ligation (NCL) of fragments corresponding to amino acids 1-33 and 34-72 was conducted. Post-ligation disulfide formation and HPLC separation gave

native M9-IL-8, misfolded isomer (Cys7-Cys50, Cys9-Cys34) and the dimer having intra (Cys34-Cys50) as well as intermolecular disulfide linkages (Figure 6A). As anticipated, the native conformer (**1**) was quiescent as a substrate of UGGT1, and misfolded isomer (**2**) and the dimer (**3**) were reactive (Figure 6B). Among non-native conformers, the dimer was clearly more reactive. The reactivity order (dimer>misfolded>folded) paralleled with their extent of fluorescent emission by the ANS binding experiment, indicating that UGGT1 indeed favors glycoproteins having surface hydrophobicity.

4.3 Glucosylation of the glycoprotein folding intermediates

Analyses of substrate recognition of UGGT1 have been carried out using model substrates which keep static state; however, actual nascent polypeptides should generate multiple folding intermediates during dynamic folding process in the ER before reaching to the native or misfolded states. More recently, our study addressed to describe the ability of UGGT1 to recognize these folding intermediates using two types of substrates generating productive folding and unproductive folding intermediates which reach the correctly folded glycoprotein and misfolded glycoprotein, respectively (Figure 7A) [36]. We used synthetic crambin carrying Man9 that is a 46 amino acid protein having three disulfide bonds. This glycosylated crambin was found to be a suitable substrate, because it generates the correctly folded state as the sole product even in the absence of denaturant such as guanidine or urea that are usually used in folding experiments. In addition to the glycosylated crambin with intact amino acid sequence (peak 1 in Figure 7B), we prepared a mutant (peak 2 in Figure 7B), wherein two of the six native cysteine residues were substituted with serine, to generate unproductive folding intermediates which is depicted as three blue colored peaks

in HPLC chromatogram of Figure 7B. Monitoring of both the glucosylation by UGGT1 and folding process of a mixture of intact and mutant glycosylated crambin as a model of a mixture of correctly folded, folding intermediates, and misfolded glycoproteins revealed that UGGT1 can glucosylate all intermediates in their folding process, indicating that UGGT1 inspects not only final folded products, but also the glycoprotein folding intermediates.

5. Enzymatic activity of UGGT isoforms

Recently, occurrence of UGGT2, an isoform of well-documented UGGT1, has been revealed in several animals. Human UGGT1 and UGGT2 are approximately 55% homologous in their amino acid sequence. The similarity is particularly high (over 80%) in the C-terminal catalytic regions. In spite of that, UGGT2 has been considered to be inactive as a glycosyltransferase, as far as tested by conventional assay system established for UGGT1. Up to now, UGGT2 has been identified from several animals [37]. On the other hand, several lines of evidence implicate its biological significance. For instance, in *C. elegans*, UGGT2 depletion study by Buzzi et al. indicated its protecting role under ER stress conditions and suggested distinct biological functions of two isoforms [38]. More recently, the effect of progesterone (P4) on the expression and activity of UGGTs were reported by Prados et al, who deductively concluded that both UGGT1 and UGGT2 are enzymatically active [39]. That C-terminal region of UGGT2 is catalytically competent was previously shown by Arnold and Kaufman. They revealed that a chimeric proteins in which

the catalytic domains of UGGT1 was replaced with that of UGGT2 was active, indicating that the catalytic domain of UGGT2 is fully functional [40].

Despite these, all the evidence from in vitro experiments argued against enzymatic activity of UGGT2. However, if UGGT1 and UGGT2 act on a different set of glycoproteins, there is a possibility that conventional assay systems, which have been optimized for UGGT1, might not be suitable for UGGT2. Indeed, our analysis that employed recombinant UGGT2 and synthetic acceptor substrate M9-BODIPY revealed the enzymatic activity of UGGT2 [41]. Substrate specificity of UGGT2 was shown to be quite similar to UGGT1 in several respects. Firstly, M8A, M7AB, M7AC, G1M9 and G2M9 were all inert, indicating that a Glc residue is solely transferred to the trimannose unit of the A-arm (Figure 8A). In addition, reactivity of glycans gradually diminished as the number of mannose residues decreased. Finally, UGGT2 behaved similarly to UGGT1 toward synthetic glycoproteins. Namely, when IL8(M9) was tested, UGGT2 most efficiently acted on the homodimer (**3**), while the native conformer was inactive. The reactivity order (**3** > **2** > **1**) was identical between two UGGT isoforms (Figure 8B). All these results suggest that UGGT2, as well as UGGT1, might play a role as a folding sensor enzyme in the ER.

In the ER, both UGGT1 and UGGT2 associate with Sep15, a 15 kDa protein which contains selenocystein (Sec) [42, 43]. Sep15 is expected to play a role in protein folding, as it contains a redox active triad, Cys-X-Sec triad [44]. Noticeably, its expression is up-regulated by ER stress [45]. Coexpression of Sep15, in which the Sec was replaced with Cys, was shown to enhance the enzymatic activities of both UGGTs. Especially, the activity of UGGT2 was dramatically increased enhanced (Figure 8C). Co-expression of

SelM, a close homolog of Sep15, which lacks the ability to associate with UGGTs, did not cause any effect.

6. Perspective

Because of the lack of substrates suitable to precise analysis, UGGT's ability to discriminate folding states of glycoproteins has been poorly addressed. Our study, which employed structurally defined substrates, provided several findings valuable in understanding its unique specificity. We believe our results indicate the strength of research strategy based on synthetic chemistry, in order for the advancement of glycobiology.

Biochemical analysis of the properties of UGGTs using synthetic oligosaccharides has provided valuable information that could not be obtained with assays using naturally occurring glycoproteins. However, the function of each domain in the UGGTs remains uncertain. In particular, the N-terminal regions of the UGGTs account for approximately 80% of the full length and are assumed to play a role in recognition of misfolded proteins; however, direct evidence of this is still lacking. To obtain detailed information about the substrate recognition mechanism of UGGTs, designing complex aglycons that more closely mimic the misfolding status of glycoproteins would be required. Furthermore, to gain an understanding of the unique characteristics of UGGTs which act as folding sensors in the ER quality control system, identification of the domains involved in aglycon recognition using chemical labelling methods should be performed.

Acknowledgment

Our research summarized here was conducted mainly at the RIKEN synthetic cellular chemistry laboratory and the JST ERATO Ito glycotriology project. We thank all the members for their excellent contribution to this project, in particular, S. Dedola, K. Fujikawa, M. Hachisu, A. Kanamori, A. Koizumi, Y. Makimura, I. Matsuo, A. Miyagawa, K. Ohara, M. Sakono, M. Takatani, K. Totani and N. Wang. We are also grateful to S. Shirahata and A. Takahashi for technical assistance.

- [1] Brodsky JL, Skach WR. Protein folding and quality control in the endoplasmic reticulum: Recent lessons from yeast and mammalian cell systems. *Curr Opin Cell Biol* 2011;23:464-75.
- [2] Hubbard SC, Robbins PW. Synthesis and processing of protein-linked oligosaccharides in vivo. *J Biol Chem* 1979;254:4568-76.
- [3] Kornfeld R, Kornfeld S. Assembly of asparagine-linked oligosaccharides. *Annu Rev Biochem* 1985;54:631-64.
- [4] Daniels R, Kurowski B, Johnson AE, Hebert DN. N-linked glycans direct the cotranslational folding pathway of influenza hemagglutinin. *Mol Cell* 2003;11:79-90.
- [5] Whitley P, Nilsson IM, von Heijne G. A nascent secretory protein may traverse the ribosome/endoplasmic reticulum translocase complex as an extended chain. *J Biol Chem* 1996;271:6241-4.
- [6] Ruiz-Canada C, Kelleher DJ, Gilmore R. Cotranslational and posttranslational N-glycosylation of polypeptides by distinct mammalian OST isoforms. *Cell* 2009;136:272-83.
- [7] Bause E. Structural requirements of N-glycosylation of proteins. Studies with proline peptides as conformational probes. *Biochem J* 1983;209:331-6.
- [8] Rutkevich LA, Williams DB. Participation of lectin chaperones and thiol oxidoreductases in protein folding within the endoplasmic reticulum. *Curr Opin Cell Biol* 2011;23:157-66.
- [9] D'Alessio C, Caramelo JJ, Parodi AJ. UDP-Glc:glycoprotein glucosyltransferase-glucosidase II, the ying-yang of the ER quality control. *Semin Cell Dev Biol* 2010;21:491-9.
- [10] Parodi AJ. Protein glucosylation and its role in protein folding. *Annu Rev Biochem* 2000;69:69-93.
- [11] Zuber C, Fan JY, Guhl B, Parodi A, Fessler JH, Parker C, et al. Immunolocalization of UDP-glucose:glycoprotein glucosyltransferase indicates involvement of pre-Golgi intermediates in protein quality control. *Proc Natl Acad Sci U S A* 2001;98:10710-5.

- [12] Sousa MC, Ferrero-Garcia MA, Parodi AJ. Recognition of the oligosaccharide and protein moieties of glycoproteins by the UDP-Glc:glycoprotein glucosyltransferase. *Biochemistry* 1992;31:97-105.
- [13] Rawitch AB, Pollock HG, Yang SX. Thyroglobulin glycosylation: location and nature of the N-linked oligosaccharide units in bovine thyroglobulin. *Arch Biochem Biophys* 1993;300:271-9.
- [14] Dorland L, van Halbeek H, Vleigenthart JF, Lis H, Sharon N. Primary structure of the carbohydrate chain of soybean agglutinin. A reinvestigation by high resolution ¹H NMR spectroscopy. *J Biol Chem* 1981;256:7708-11.
- [15] Fu D, Chen L, O'Neill RA. A detailed structural characterization of ribonuclease B oligosaccharides by ¹H NMR spectroscopy and mass spectrometry. *Carbohydr Res* 1994;261:173-86.
- [16] Ritter C, Helenius A. Recognition of local glycoprotein misfolding by the ER folding sensor UDP-glucose:glycoprotein glucosyltransferase. *Nat Struct Biol* 2000;7:278-80.
- [17] Caramelo JJ, Castro OA, Alonso LG, De Prat-Gay G, Parodi AJ. UDP-Glc:glycoprotein glucosyltransferase recognizes structured and solvent accessible hydrophobic patches in molten globule-like folding intermediates. *Proc Natl Acad Sci U S A* 2003;100:86-91.
- [18] Caramelo JJ, Castro OA, de Prat-Gay G, Parodi AJ. The endoplasmic reticulum glucosyltransferase recognizes nearly native glycoprotein folding intermediates. *J Biol Chem* 2004;279:46280-5.
- [19] Molinari M. N-glycan structure dictates extension of protein folding or onset of disposal. *Nat Chem Biol* 2007;3:313-20.
- [20] Sousa M, Parodi AJ. The molecular basis for the recognition of misfolded glycoproteins by the UDP-Glc:glycoprotein glucosyltransferase. *EMBO J* 1995;14:4196-203.
- [21] Trombetta ES, Helenius A. Conformational requirements for glycoprotein reglucosylation in the endoplasmic reticulum. *J Cell Biol* 2000;148:1123-29.
- [22] Ritter C, Quirin K, Kowarik M, Helenius A. Minor folding defects trigger local modification of glycoproteins by the ER folding sensor GT. *EMBO J* 2005;24:1730-8.
- [23] Matsuo I, Totani K, Tatami A, Ito Y. Comprehensive synthesis of ER related high-mannose-type sugar chains by convergent strategy. *Tetrahedron* 2006;62:8262-77.
- [24] Matsuo I, Wada M, Manabe S, Yamaguchi Y, Otake K, Kato K, et al. Synthesis of monoglucosylated high-mannose-type dodecasaccharide, a putative ligand for molecular chaperone, calnexin, and calreticulin. *J Am Chem Soc* 2003;125:3402-3.
- [25] Totani K, Ihara Y, Matsuo I, Koshino H, Ito Y. Synthetic substrates for an endoplasmic reticulum protein-folding sensor, UDP-glucose: glycoprotein glucosyltransferase. *Angew Chem Int Ed* 2005;44:7950-4.
- [26] Totani K, Matsuo I, Ihara Y, Ito Y. High-mannose-type glycan modifications of dihydrofolate reductase using glycan-methotrexate conjugates. *Bioorg Med Chem* 2006;14:5220-9.
- [27] Koizumi A, Matsuo I, Takatani M, Seko A, Hachisu M, Takeda Y, et al. Top-down chemoenzymatic approach to a high-mannose-type glycan library: synthesis of a common precursor and its enzymatic trimming. *Angew Chem Int Ed* 2013;52:7426-31.

- [28] Trombetta SE, Bosch M, Parodi AJ. Glucosylation of glycoproteins by mammalian, plant, fungal, and trypanosomatid protozoa microsomal membranes. *Biochemistry* 1989;28:8108-16.
- [29] Keith N, Parodi AJ, Caramelo JJ. Glycoprotein tertiary and quaternary structures are monitored by the same quality control mechanism. *J Biol Chem* 2005;280:18138-41.
- [30] Zapun A, Darby NJ, Tessier DC, Michalak M, Bergeron JJ, Thomas DY. Enhanced catalysis of ribonuclease B folding by the interaction of calnexin or calreticulin with ERp57. *J Biol Chem* 1998;273:6009-12.
- [31] Totani K, Ihara Y, Tsujimoto T, Matsuo I, Ito Y. The recognition motif of the glycoprotein-folding sensor enzyme UDP-Glc:glycoprotein glucosyltransferase. *Biochemistry* 2009;48:2933-40.
- [32] Miyagawa A, Totani K, Matsuo I, Ito Y. Promiscuous activity of ER glucosidase II discovered through donor specificity analysis of UGGT. *Biochem Biophys Res Commun* 2010;403:322-8.
- [33] Sakono M, Seko A, Takeda Y, Hachisu M, Ito Y. Biophysical properties of UDP-glucose:glycoprotein glucosyltransferase, a folding sensor enzyme in the ER, delineated by synthetic probes. *Biochem Biophys Res Commun* 2012;426:504-10.
- [34] Unverzagt C, Kajihara Y. Chemical assembly of N-glycoproteins: a refined toolbox to address a ubiquitous posttranslational modification. *Chem Soc Rev* 2013;42:4408-20.
- [35] Izumi M, Makimura Y, Dedola S, Seko A, Kanamori A, Sakono M, et al. Chemical synthesis of intentionally misfolded homogeneous glycoprotein: a unique approach for the study of glycoprotein quality control. *J Am Chem Soc* 2012;134:7238-41.
- [36] Dedola S, Izumi M, Makimura Y, Seko A, Kanamori A, Sakono M, et al. Folding of synthetic homogeneous glycoproteins in the presence of a glycoprotein folding sensor enzyme. *Angew Chem Int Ed* 2014;53:2883-7.
- [37] Arnold SM, Fessler LI, Fessler JH, Kaufman RJ. Two homologues encoding human UDP-glucose:glycoprotein glucosyltransferase differ in mRNA expression and enzymatic activity. *Biochemistry* 2000;39:2149-63.
- [38] Buzzi LI, Simonetta SH, Parodi AJ, Castro OA. The two *Caenorhabditis elegans* UDP-glucose:glycoprotein glucosyltransferase homologues have distinct biological functions. *PLoS One* 2011;6:e27025.
- [39] Prados MB, Caramelo JJ, Miranda SE. Progesterone regulates the expression and activity of two mouse isoforms of the glycoprotein folding sensor UDP-Glc: Glycoprotein glucosyltransferase (UGGT). *Biochim Biophys Acta* 2013;1833:3368-74.
- [40] Arnold SM, Kaufman RJ. The noncatalytic portion of human UDP-glucose: glycoprotein glucosyltransferase I confers UDP-glucose binding and transferase function to the catalytic domain. *J Biol Chem* 2003;278:43320-8.
- [41] Takeda Y, Seko A, Hachisu M, Daikoku S, Izumi M, Koizumi A, et al. Both isoforms of human UDP-glucose:glycoprotein glucosyltransferase are enzymatically active. *Glycobiology* 2014;24:344-50.
- [42] Korotkov KV, Kumaraswamy E, Zhou Y, Hatfield DL, Gladyshev VN. Association between the 15-kDa selenoprotein and UDP-glucose:glycoprotein glucosyltransferase in the endoplasmic reticulum of mammalian cells. *J Biol Chem* 2001;276:15330-6.

- [43] Labunskyy VM, Ferguson AD, Fomenko DE, Chelliah Y, Hatfield DL, Gladyshev VN. A novel cysteine-rich domain of Sep15 mediates the interaction with UDP-glucose:glycoprotein glucosyltransferase. *J Biol Chem* 2005;280:37839-45.
- [44] Ferguson AD, Labunskyy VM, Fomenko DE, Arac D, Chelliah Y, Amezcua CA, et al. NMR structures of the selenoproteins Sep15 and SelM reveal redox activity of a new thioredoxin-like family. *J Biol Chem* 2006;281:3536-43.
- [45] Labunskyy VM, Yoo MH, Hatfield DL, Gladyshev VN. Sep15, a thioredoxin-like selenoprotein, is involved in the unfolded protein response and differentially regulated by adaptive and acute ER stresses. *Biochemistry* 2009;48:8458-65.

Figure captions

Figure 1. (A) Overview of glycoprotein folding and quality control process in the ER. (B) Schematic representation of UGGT1.

Figure 2. Systematic synthesis of high-mannose-type oligosaccharides. (A) Convergent synthesis of high-mannose-type oligosaccharide. (B) Synthetic N-glycan molecular probes by introducing methotrexate (MTX). The term "CHO" means carbohydrate. MTX bind tightly with DHFR to form glycan-attached proteins (CHO-MTX-DHFR).

Figure 3. Schematic illustration of Top-down chemo-enzymatic approach to construct high-mannose-type glycan library. Each glycan could be obtained from a synthetic common precursor by orthogonal enzymatic trimming.

Figure 4. Analysis of UGGT. (A) UGGT activity against synthetic M9-derivatives. Enzymatic reaction was carried out in 4 mM Tris-HCl (pH 8.0), 10 mM CaCl₂, 50 μM deoxynojirimycin, and 0.6% Triton X-100, containing rat liver UGGT (80 ng/μL), UDP-[³H]-Glc (0.4 mCi), and 50 μM of M9-derivatives. After incubation at 37 °C for 60 min, the reaction was quenched and incorporated radioactivity was analyzed by liquid scintillation counting. (B) Glycan specificity analysis of UGGT and kinetics of CHO-MTX. Enzymatic reaction was carried out in 4 mM Tris-HCl (pH 8.0), 10 mM CaCl₂, 0.6 mM deoxynojirimycin, 0.6 mM deoxymannojirimycin, and 0.6% Triton X-100, containing rat

liver UGGT (4 μ g), 10 mM UDP-Glc, and 100 μ M of CHO-MTX at 37 °C. Each initial velocity was calculated from the Glc transfer at 1 h incubation.

Figure 5. Analyses of substrate recognition of UGGT using synthetic substrates. (A) The initial velocity of the PEG-linked M9-BODIPY substrates. Enzymatic reaction was carried out in 4 mM Tris-HCl (pH 8.0), 10 mM CaCl₂, 0.25 mM deoxynojirimycin, 0.25 mM deoxymannojirimycin, and 0.6% Triton X-100, containing rat liver UGGT (5 μ g/ μ L), 2.5 mM UDP-Glc, and 25 μ M of M9-(PEG0-12)-BODIPY at 37 °C. Each initial velocity was calculated from the Glc transfer at 1 h incubation. (B) CD spectra of recombinant *Drosophila* UGGT1 expressed in *Escherichia coli* (0.3 μ M) with various concentration of M9-G-E-5-TAMRA. The concentration of substrate was adjusted from 0 to 20 μ M.

Figure 6. (A) Schematic representation of native and misfolded M9-IL-8s. (B) Percentage of consumed M9-IL-8 (**1**, **2**, and **3**) by UGGT1 reaction after 2 h (green), 6 h (orange), and 20 h (blue). Reaction condition was following: 50 mM Tris-HCl (pH 7.5), 5 mM CaCl₂, recombinant human UGGT1 expressed in *Escherichia coli* (0.6 ng/ μ L), 0.5 mM UDP-Glc, and 4 μ M of M9-IL-8 at 37 °C.

Figure 7. (A) UGGT1 inspection during the glycoprotein folding process. Native glycoproteins can follow a productive folding course to yield correctly folded glycoproteins, or an unproductive folding course, thus providing misfolded glycoprotein. (B) LC-UV profile at different time points of folding process for the mixture of native sequence (peak 1,

productive folding) and mutated sequence (peak 2, unproductive folding) of M9-carrying crambin. Reaction condition was following: 50 mM Tris-HCl (pH 7.5), 5 mM CaCl₂, recombinant human UGGT1 expressed in *Escherichia coli* (4.5 ng/μL), 0.5 mM UDP-Glc, and 7.5 μM of M9-carrying crambin at 37 °C.

Figure 8. (A) Glycoform specificity of UGGT2 which was compared by quantification of the Glc transfer to BODIPY labeled high-mannose-type glycans. Reaction condition was following: 4 mM Tris-HCl (pH 7.6), 2 mM CaCl₂, 0.05% Triton X-100, 1 mM 2-mercaptoethanol, recombinant human UGGT1 expressed in 293T cells, 1 mM UDP-Glc, and 80 nM M9-BODIPY at 37 °C for 6 h. (B) Percentage of consumed M9-IL-8 (**1**, **2**, and **3**) by the reaction of UGGT2 after 22 h (orange) and 42 h (blue). Reaction condition was following: 16 mM Tris-HCl (pH 7.3), 4 mM CaCl₂, recombinant human UGGT1 expressed in 293T human embryonic kidney cells, 1 mM UDP-Glc, and 5 μM of M9-IL-8 at 37 °C. (C) Comparison of glucosyltransferase activities of UGGT2 co-expressed with Sep15 or SelM.

Fig. 1

A

B

Fig. 2

Fig. 3

A

B

	K_m [μM]	V_{max} [nmol/h/mg]
M9-MTX	207	32.3
M8B-MTX	197	15.4
M8C-MTX	448	23.8
M7-MTX	46	4.8

Fig. 4

Fig. 5

A

B

A **B** **Fig. 7**

Fig. 8