BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

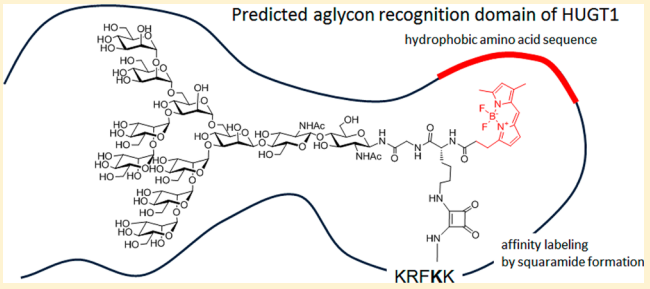
Profiling Aglycon-Recognizing Sites of UDP-glucose:glycoprotein Glucosyltransferase by Means of Squarate-Mediated Labeling

Keiichiro Ohara,[†] Yoichi Takeda,[†,§] Shusaku Daikoku,[†] Masakazu Hachisu,[†,∥] Akira Seko,[†] and Yukishige Ito*[,†,‡]

[†]Japan Science and Technology Agency (JST), ERATO, Ito Glycotrilogy Project, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan
[‡]RIKEN, Synthetic Cellular Chemistry Laboratory, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan

Supporting Information

ABSTRACT: Because of its ability to selectively glucosylate misfolded glycoproteins, UDP-glucose:glycoprotein glucosyltransferase (UGGT) functions as a folding sensor in the glycoprotein quality control system in the endoplasmic reticulum (ER). The unique property of UGGT derives from its ability to transfer a glucose residue to N-glycan moieties of incompletely folded glycoproteins. We have previously discovered nonproteinic synthetic substrates of this enzyme, allowing us to conduct its high-sensitivity assay in a quantitative manner. In this study, we aimed to conduct site-selective affinity labeling of UGGT using a functionalized oligosaccharide probe to identify domain(s) responsible for recognition of the aglycon moiety of substrates. To this end, a probe **1** was designed to selectively label nucleophilic amino acid residues in the proximity of the canonical aglycon-recognizing site of human UGGT1 (HUGT1) via squaramide formation. As expected, probe **1** was able to label HUGT1 in the presence of UDP. Analysis by nano-LC-ESI/MS^n identified a unique lysine residue (K1424) that was modified by **1**. Kyte−Doolittle analysis as well as homology modeling revealed a cluster of hydrophobic amino acids that may be functional in the folding sensing mechanism of HUGT1.



A majority of nascent polypeptides that enter the secretory pathway are subject to N-glycosylation, which occurs in the lumen of the endoplasmic reticulum (ER). This modification occurs co-translationally[1] or post-translationally[2] by *en bloc* transfer of a $Glc_3Man_9GlcNAc_2$ oligosaccharide bound to dolichol pyrophosphate to Asn residues embedded in a Asn-Xaa-Ser/Thr consensus triad (where Xaa is any amino acid except Pro)[3] through the action of oligosaccharyltransferase. The outermost and penultimate glucose residues are successively removed by sequential action of glucosidase I (G-I) and glucosidase II (G-II) to generate a monoglucosylated glycoform $GlcMan_9GlcNAc_2$, a key biosynthetic intermediate in the glycoprotein folding process in the ER. Glycoproteins that have acquired this glycoform are captured by lectin-like chaperones, calnexin (CNX) and calreticulin (CRT), which share specific affinity for monoglucosylated high-mannose-type glycans. By virtue of their intrinsic chaperone activity along with association with a protein disulfide isomerase-like protein, ERp57, CNX and CRT assist in correct folding of client glycoproteins.[4]

Further trimming of glucose residue by G-II leads glycoproteins to the $Man_9GlcNAc_2$ glycoform, allowing them to dissociate from CNT or CRT.[4] Whereas glycoproteins that have achieved mature folding are exported to the Golgi apparatus, incompletely folded glycoproteins are reglucosylated by UDP-glucose:glycoprotein glucosyltransferase (UGGT), so that they can reassociate with CNT/CRT.[5] Intriguingly, UGGT transfers a glucose residue predominately to the glycoprotein that displays a non-native conformation such as a molten globule-like folding intermediate, while both correctly folded and irreparably misfolded glycoproteins are poorly glucosylated by UGGT.[5,6] Namely, in the N-glycan-dependent protein folding process called the CNT/CRT cycle,[4] UGGT functions as the folding sensor, as it can survey folding state of glycoproteins.[5]

UGGT is a soluble glycoprotein of approximately 160 kDa, which is widespread in various organisms.[7] The enzyme exists in a monomeric form composed of at least two domains, the N-terminal domain that comprises 80% of the protein and the C-terminal domain that corresponds to the rest of the molecule.[8] Whereas the former does not show any similarity to other proteins, the latter has considerable similarity to members of glycosyltransferase family 8. Consequently, it has been suggested that the C-terminal domain is responsible for the glucosyltransferase activity, while the larger N-terminal domain participates in conformation sensing.[5] To understand the unique specificity of UGGT, a large body of work in which denatured glycoproteins were used as model substrates has been conducted.[9−12] Although these studies have revealed certain properties common to substrates of UGGT, an unambiguous

Received: April 30, 2015
Published: July 21, 2015

DOI: 10.1021/acs.biochem.5b00785
Biochemistry 2015, 54, 4909−4917

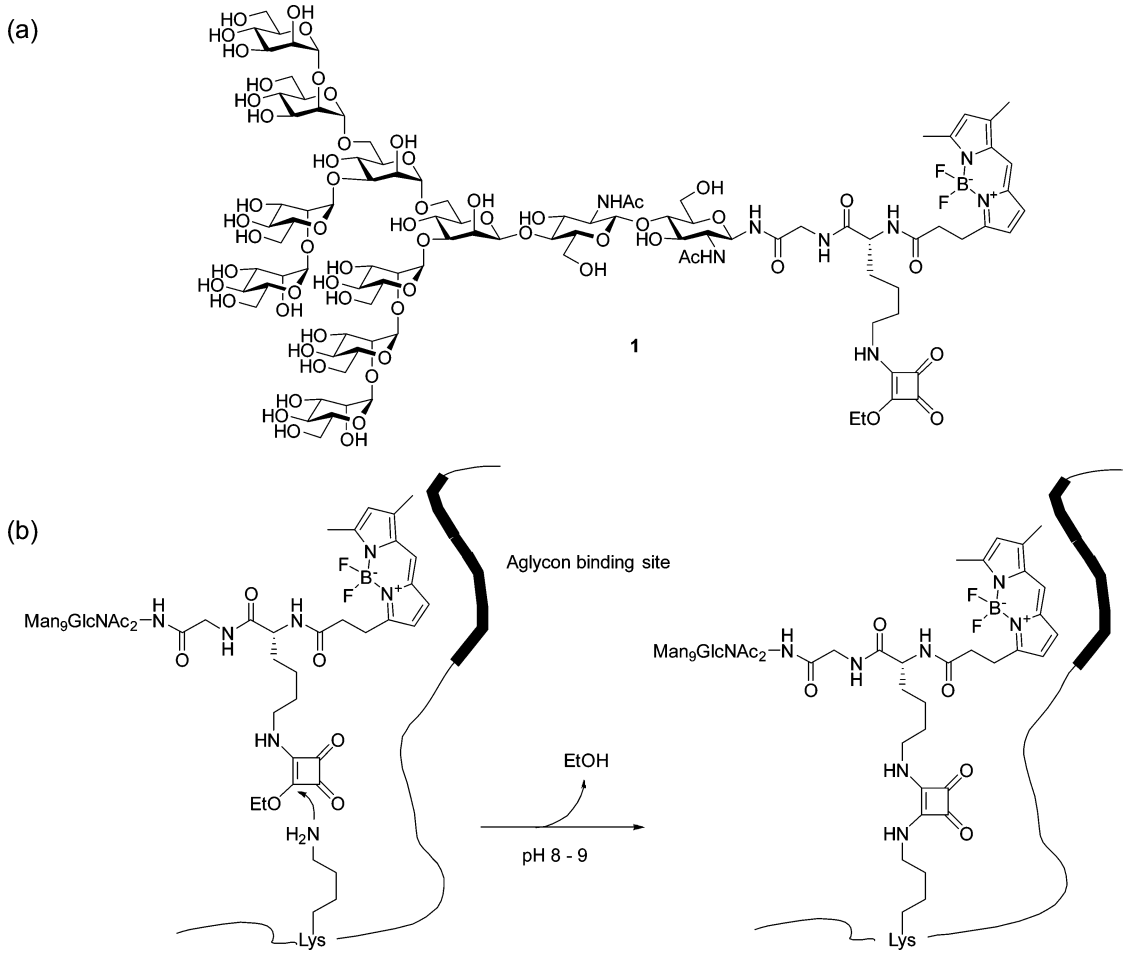


Figure 1. (a) Structure of probe **1** and (b) labeling of Lys proximal to the canonical aglycon binding site.

conclusion could not be drawn because of the intrinsic heterogeneity of glycoproteins.

Addressing these problems, we previously developed nonproteinic substrates of UGGT, composed of high-mannose-type oligosaccharide and aglycon with chromophoric or fluorophoric properties.[13−15] Utilization of these substrates allowed us not only to unambiguously determine glycan structures recognized by UGGT[13] but also to reveal the glucosyltransferase activity of UGGT2,[16] an isoform of UGGT that had been believed to be enzymatically inactive.[17,18] In these studies, we have also revealed the influences of aglycon structure on the activity of synthetic substrates.[14]

In these cases, it was assumed that the aglycon moieties of the glycoconjugates mimic the hydrophobic motif of unfolded proteins. Indeed, glycopeptides with a cluster of hydrophobic amino acids were shown to be reactive substrates of UGGT.[6,9] This speculation was supported by a recent study by Izumi et al.,[19] which employed chemically synthesized glycoproteins as substrates of UGGT. The study allowed creation of homogeneous glycoproteins in folded as well as intentionally misfolded forms and clarified that their surface hydrophobicity parallels the reactivity toward UGGT. Although efforts have been made to elucidate the characteristic properties of UGGT, how UGGT can recognize the hydrophobic moiety of its substrate has been poorly understood. While it has been suggested that the 80% N-terminal domain is involved in the recognition of the hydrophobic surface of misfolded proteins,[18] a possibility that catalytic as well as folding recognition sites reside in the C-terminal domain has not been ruled out.[8]

This study aimed to localize the conformation-sensing region of UGGT using a nonproteinic synthetic substrate. For this purpose, we planned to conduct site-selective labeling using an affinity probe, an approach often used to seek ligand binding sites of proteins,[20] especially when ligand-interacting proteins are large, as in the case of UGGT, whose structural characterization by X-ray or nuclear magnetic resonance (NMR) is difficult. Usually, execution of this approach calls for a strategy that allows covalent linkage formation with nucleophilic amino acid residues, typically lysine[21] or cysteine,[22] proximal to ligand binding site. Namely, labeling with an analogue bearing a reactive electrophile[20] is followed by characterization of tryptic peptide fragments of labeled protein. For selective labeling of the aglycon binding sites, it would be most pertinent to target amino groups of lysine residues, as they are far more abundant than cysteine. In this context, we focused our attention on using a squarate group as an electrophile, which is known to be reactive toward amine yet stable in aqueous environments.[23−25] As BODIPY-modified $Man_9GlcNAc_2$ derivatives with various spacer lengths were shown to be excellent substrates of UGGT,[13] an oligosaccharide probe **1** was designed, which was equipped with a squaric amide ester as a reactive group (Figure 1).

MATERIALS AND METHODS

All chemical reagents were purchased from commercial sources and used without further purification. Preparative high-perform-

DOI: 10.1021/acs.biochem.5b00785
Biochemistry 2015, 54, 4909−4917

ance liquid chromatography (HPLC) was conducted on JASCO PU-2087 and UV-2070 instruments with an InertSustain C18 column (5 μm particle size, 14 mm × 250 mm, GL Science). Analytical HPLC was performed on Waters e2695 and 2475 instruments with an InertSustain C18 column (3 μm particle size, 0.46 cm × 25 cm, GL Science) or a C18 Mightysil column (5 μm particle size, 0.46 cm × 25 cm, KANTO). ^{1}H NMR spectra were recorded employing a Bruker Avance 500 spectrometer at 25 °C. Chemical shifts were referenced to HCD_2OD [δ 3.31 in CD_3OD/D_2O (4/1, v/v)], HOD (δ 4.79 in D_2O), or HCD_2CN [δ 1.94 in CD_3CN/D_2O (1/1, v/v)]. Coupling constants are given in hertz. Signal splitting patterns are described as singlet (s), doublet (d), triplet (t), double of doublet (dd), multiplet (m), and broad signal (brs). MALDI-TOF mass spectra (MALDI-TOF MS) were recorded on a Bruker Daltonics autoflex speed instrument using SDHB (9/1 mixture of 2,5-dihydroxybenzoic acid and 2-hydroxy-5-methoxybenzoic acid) as a matrix. Sequencing grade modified trypsin was purchased from Promega.

Expression of Human UGGT1 (HUGT1) Is Truncated in 293T Cells. A cDNA encoding HUGT1 (38−744 amino acids) with a stop codon at the 3′-terminus was inserted between NotI and XbaI sites of p3XFLAG-CMV-9 (Sigma-Aldrich, St. Louis, MO), which is constructed for the attachment of an N-terminal FLAG tag. Similarly, a cDNA encoding HUGT1 (745−1555 amino acids) or its full-length form (38−1555 amino acids) was inserted into the same plasmid. Expression of these recombinant proteins and their purification using anti-FLAG M2 agarose beads (Sigma-Aldrich) were performed as described previously.[15]

Labeling of HUGT1 and Truncates. The following solutions were prepared. Solution 1 contained 55.6 mM Tris-HCl (pH 8.0), 1.1 mM $CaCl_2$, 1.1 mM 2-mercaptoethanol, and 1.1 mM UDP-Glc. Solution 2 contained 55.6 mM Tris-HCl (pH 8.0), 1.1 mM $CaCl_2$, 1.1 mM 2-mercaptoethanol, 1.1 mM UDP-Glc, and 1.1 μM probe **1**. Solution 3 contained 55.6 mM Tris-HCl (pH 8.0), 1.1 mM $CaCl_2$, 1.1 mM 2-mercaptoethanol, and 1.1 μM probe **1**. Solution 4 contained 55.6 mM Tris-HCl (pH 8.0), 1.1 mM $CaCl_2$, 1.1 mM 2-mercaptoethanol, 1.1 mM UDP, and 1.1 μM probe **1**. To 9 μL of each of these solutions was added 1 μL of an ≈50% slurry suspension of M2 agarose beads containing ≈0.37 μg of immobilized HUGT1 (or truncated HUGT1), 20 mM HEPES-NaOH (pH 7.2), and 10% glycerol. The samples were incubated at 37 °C for 12 h in the dark before 1 μL of a 100 mM lysine solution in 40 mM Tris-HCl (pH 8.0) was added. After incubation at 37 °C for 2 h, 4 μL of LDS sample buffer (Life Technologies) containing 10 mM 2-mercaptoethanol was added. The samples were heated at 100 °C for 5 min and then separated on a NuPAGE 4 to 12% Bis-Tris gel (Life Technologies) using MOPS running buffer (Life Technologies) by electrophoresis (run for 50 min at 200 V). The gel was rinsed with deionized water over 2 h and then laser-scanned with the Molecular Imager FX system (Bio-Rad) using an excitation/emission filter set for Alexa488. After a fluorescence image had been recorded, the gel was stained with Quick-CBB PLUS (Wako) and photographed with an ImageQuant LAS-4000 system (GE Healthcare).

Preparation of Labeled HUGT1 for Tryptic Digestion. A solution containing 55.6 mM Tris-HCl (pH 8.0), 1.1 mM $CaCl_2$, 1.1 mM 2-mercaptoethanol, 1.1 mM UDP, and 1.1 μM probe **1** was prepared with a total volume of 450 μL and divided into 9 μL aliquots. To each solution was added 1 μL of an ≈50% slurry suspension of M2 agarose beads containing 0.37 μg of immobilized HUGT1, 20 mM HEPES-NaOH (pH 7.2), and 10% glycerol. The samples were incubated at 37 °C for 12 h before 1 μL of a 100 mM lysine solution in 40 mM Tris-HCl (pH 8.0) was added. After being incubated at 37 °C for a further 2 h, all samples were transferred to a filter, and the soluble components were removed by filtration. The beads were then washed three times with deionized water, recovered, and resuspended in deionized water (total volume of ≈100 μL).

Tryptic Digestion of Labeled HUGT1. To 70 μL of 71.4 mM Tris-HCl (pH 8.0) buffer was added 20 μL of the suspension containing labeled HUGT1, and then 10 μL of a trypsin solution (0.1 μg/μL) in resuspension buffer (Promega) was added to the suspension. After incubation at 37 °C for 3 h, the suspension was transferred to an Ultrafree centrifugal filter device (Durapore, PVDF 0.1 μm, Millipore) and centrifuged at 7500*g* for 1 min. The beads were suspended in deionized water and centrifuged at 7500*g* for 1 min. The filtrates were combined and freeze-dried.

Reverse Phase HPLC Fractionation and LC−MS Analysis of the Digested Peptide Mixture. The peptide mixture was subjected to reverse phase HPLC using a Vydac C4 reverse phase column (4.6 mm × 250 mm, 214TP54, GRACE) and eluted at a flow rate of 1.0 mL/min, first isocratically with 5% aqueous acetonitrile for 5 min, followed by a liner gradient up to 100% acetonitrile in 95 min. Fractions were collected by fluorescent emission signals at 520 nm after excitation at 488 nm and dried *in vacuo*.

The fractions were further separated using a Nano Frontier nLC digital nanoflow HPLC system (Hitachi High-Technologies Corp., Tokyo, Japan). A tapered capillary column was used as a sprayer tip and was filled with C18 silica particles [particle size of 3 μm, 75 μm (inside diameter) × 150 mm (length); Nikkyo Technos Co., Ltd., Tokyo, Japan]. Mobile phase A was 0.1% formic acid in water, whereas mobile phase B was 0.1% formic acid in acetonitrile (LC−MS grade; Merck, Darmstadt, Germany). Gradient conditions were as follows: 0 to 5 min, isocratic 10% B; 5.1 to 30 min, linear gradient from 10 to 90% B; 30.1 to 120 min, isocratic 95% B.

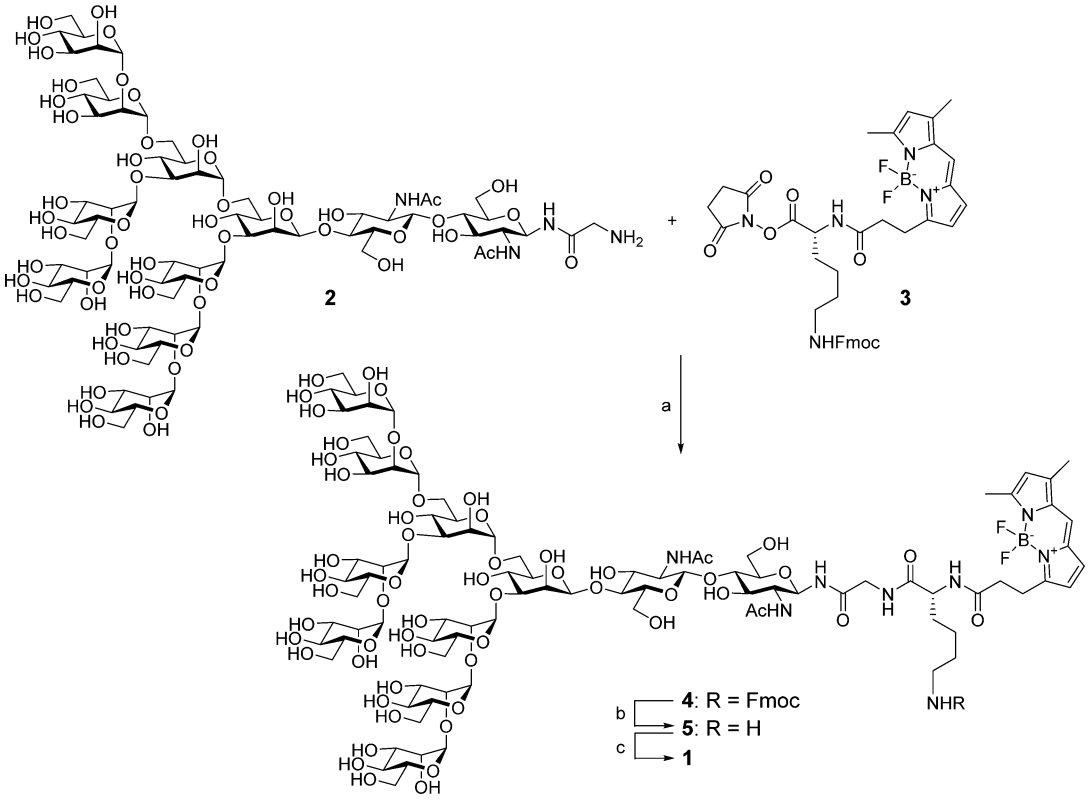
An Optitool FLE1100 Type B instrument (Ikeda Scientific Co. Ltd., Tokyo, Japan) equipped with a fiber optic cable and a collimator was used for irradiating excitation light (excitation at 470 nm) and acquiring emission light (emission at 520 nm). The collimator was fixed at a distance of 2 mm from the tapered sprayer tips and approximately 5 mm from the edge of the tip using a three-axis oil hydraulic micromanipulator. Data were collected at 200 ms intervals using Smart Chrom software (KYA Technologies Co., Tokyo, Japan).

The samples were analyzed using a quadrupole ion trap mass spectrometer (QIT-MS) coupled with a nanoelectrospray interface (amaZon ETD, Bruker Daltonics GmbH, Bremen, Germany), which is attached to a three-dimensional manipulator for fluorescence detection. The parameters for analysis were as follows: (1) dry temperature, 120 °C; (2) dry gas (N_2), 3.0 L/min; (3) scan range, m/z 600−2250; (4) compound stability, 100%; (5) target mass, m/z 1750; (6) ion charge control (ICC), on, target, 400000; (7) maximal accumulation time, 200 ms; (8) average, five spectra; and (9) polarity, positive. In MS/MS experiments, the end-cap radiofrequency amplitude was 1.2 V and the isolation width was m/z 4.0.

Kyte−Doolittle Hydropathy Analysis and Homology Modeling. Multiple-sequence alignments were performed with ClustalX 2.1.[26] Kyte−Doolittle hydropathy[27] analyses were performed using ProtScale on the ExPASy server of the Swiss Institute of Bioinformatics (http://web.expasy.org/protscale/).

DOI: 10.1021/acs.biochem.5b00785
Biochemistry 2015, 54, 4909−4917

Scheme 1. Synthesis of Probe 1[a]



[a]Reagents and conditions: (a) DIPEA, DMF, rt, 60%; (b) 5% piperidine, DMF, 0 °C, 73%; (c) diethyl squarate, DIPEA, MeOH/H_2O (1/1, v/v), room temperature, 44%.

The Phyre2 web server[28] was used for protein structure homology modeling in intensive mode (http://www.sbg.bio.ic.ac.uk/phyre2/). The Protein Data Bank (PDB) file was then visualized using the PyMOL Molecular Graphics System (DeLano Scientific, San Carlos, CA).

RESULTS

To begin, synthesis of probe **1** was conducted through formation of an amide bond between glycine-modified undecasaccharide **2**[29] and activated ester **3**[30] to afford **4** (Scheme 1). Removal of the Fmoc group gave **5**, treatment of which with excess diethyl squarate in the presence of diisopropylethylamine furnished the designed probe **1**.

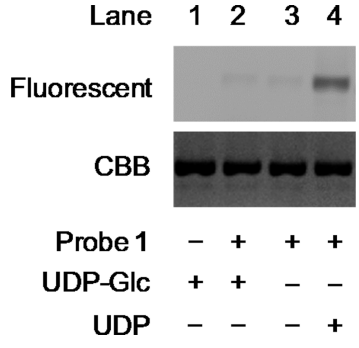


Figure 2. SDS−PAGE analysis of the labeling of HUGT1 with probe **1**: lane 1, HUGT1 incubated without a labeling agent; lane 2, HUGT1 incubated with probe **1** (1.0 μM) in the presence of UDP-Glc (1.0 mM); lane 3, HUGT1 incubated with probe **1** (1.0 μM) in the absence of UDP-Glc; lane 4, HUGT1 incubated with probe **1** (1.0 μM) in the presence of UDP (1.0 mM).

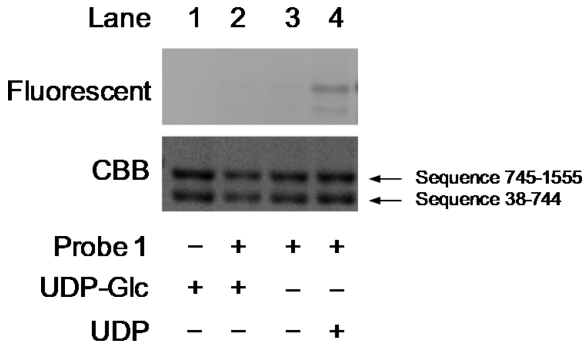


Figure 3. SDS−PAGE analysis of the labeling of the truncated HUGT1 with probe **1**: lane 1, truncated HUGT1 incubated without a labeling agent; lane 2, truncated HUGT1 incubated with probe **1** (1.0 μM) in the presence of UDP-Glc (1.0 mM); lane 3, truncated HUGT1 incubated with probe **1** (1.0 μM) in the absence of UDP-Glc; lane 4, truncated HUGT1 incubated with probe **1** (1.0 μM) in the presence of UDP (1.0 mM).

Compound **1** was then evaluated for its ability to label UGGT. As the enzyme source, we chose recombinant human UGGT1 (HUGT1) expressed in 293T human kidney cells, as it was shown to have a high enzymatic activity.[31] To begin, a mixture of HUGT1 and probe **1** was incubated in the presence or absence of 1 mM UDP-Glc and 1 mM $CaCl_2$ under slightly basic conditions (pH 8.0), which was, after incubation at 37 °C for 12 h, resolved by sodium dodecyl sulfate−polyacrylamide gel electrophoresis (SDS−PAGE). The gels were visualized by fluorescent imaging

DOI: 10.1021/acs.biochem.5b00785
Biochemistry 2015, 54, 4909−4917

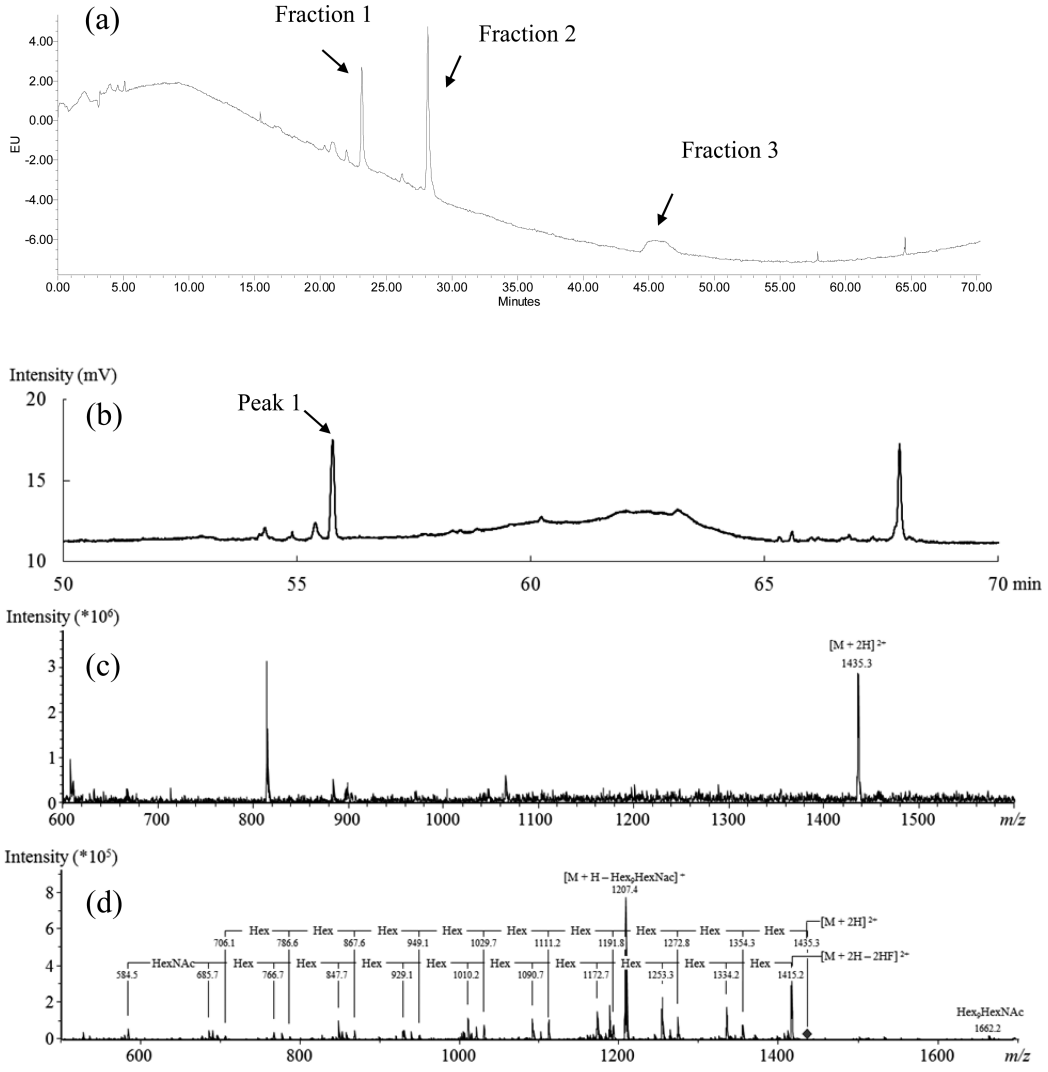


Figure 4. (a) Reverse phase (C4) fluorescent (excitation at 488, emission at 520) HPLC chromatogram of the tryptic products of labeled HUGT1. (b) Reverse phase (C18) fluorescent (excitation at 470, emission at 520) HPLC chromatogram of fraction 3. (c) ESI-MS spectrum of peak 1. (d) MS/MS spectrum of a precursor ion (m/z 1435.3).

and Coomassie Brilliant Blue (CBB) staining (Figure 2). While almost no labeling was observed in either case (lanes 2 and 3), analysis of the reaction mixture by MALDI-TOF MS revealed that probe **1** was glucosylated when UDP-Glc was included (Figure S1 of the Supporting Information). This result indicates that the compound was recognized as a substrate of HUGT1. On the other hand, HUGT1 was strongly labeled when the incubation was conducted in the presence of 1 mM UDP (Figure 2, lane 4). Under the same conditions, the protein was not labeled with oligosaccharide **5** lacking the squarate group (Figure S2 of the Supporting Information). These results strongly suggest that HUGT1 was labeled via squaramide formation. Because probe **1** scarcely reacted with free lysine even at high concentration (1 μM probe **1** with 1 mM lysine) (Figure S3 of the Supporting Information), smooth formation of the squaramide between HUGT1 and probe **1** was attributed to a proximity effect caused by their association.

The C-terminal catalytic region of UGGT belongs to glycosyltransferase (GT) family 8 and possesses two Asp-X-Asp (DXD) triads, which are found in many GTs.[32] These enzymes transfer sugar residues from the corresponding sugar nucleotide with retention of configuration. To explain the stereochemical outcome, a double-displacement mechanism via a covalently bound glycosyl–enzyme intermediate is widely accepted, in which a leaving nucleoside diphosphate is bound to the enzyme by coordinating a divalent metal cation with a carboxylate group of the DXD motif, while its diphosphate moiety acts as a base catalyst to activate an incoming hydroxyl group of an acceptor.[32] Hence, it is likely that the binding of probe **1** to HUGT1 was enhanced by the complexation of UDP with the enzyme active site.

We subsequently conducted the labeling experiments using FLAG-tagged fragments 38–744 and 745–1555, the former corresponding to the 50% N-terminal domain of HUGT1 and the latter to the 50% C-terminal domain. A mixture of the polypeptides obtained by co-expression of these sequences showed an enzymatic activity comparable with that of the full-length enzyme, as reported by Guerin and Parodi[8] for *Schizosaccharomyces pombe* UGGT. It was assumed that these polypeptides associate noncovalently to form an enzymatically active complex. As anticipated, fluorescent labeling was observed only in the presence of UDP (Figure 3, lane 4). Noticeably, the sequence of residues 744–1555 was more strongly labeled than the sequence of residues 38–744, indicating that the labeling

DOI: 10.1021/acs.biochem.5b00785
Biochemistry 2015, 54, 4909–4917

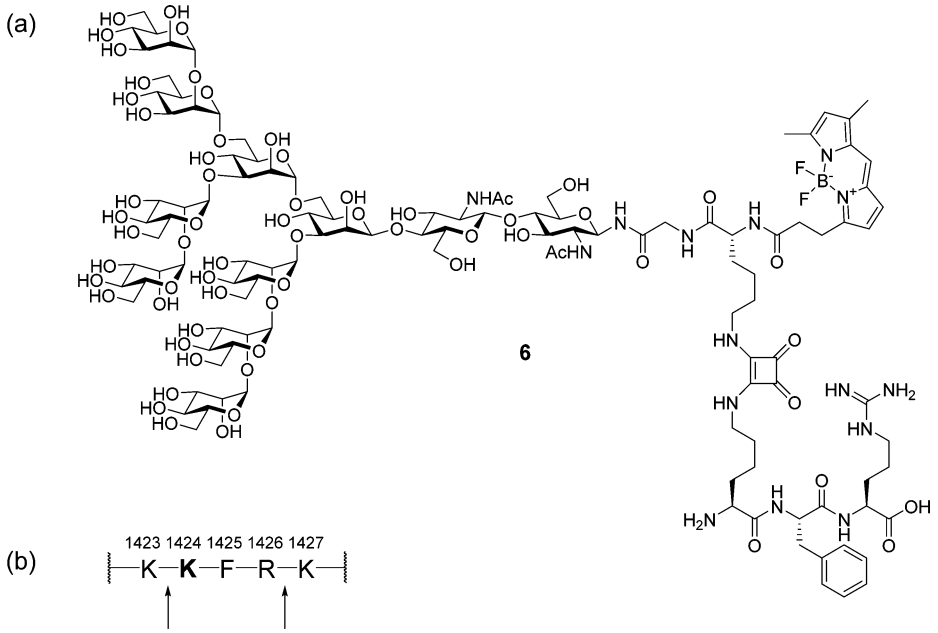


Figure 5. (a) Structure of glycopeptide **6**. (b) Labeling and tryptic cleavage sites of HUGT1. The labeled lysine residue is indicated in bold, and the arrows indicate the tryptic cleavage points.

occurred predominantly toward the C-terminal half of the molecule.

To determine the site that was covalently modified by **1**, labeled HUGT1 was digested with trypsin and the resulting fragments were fractionated on a reverse phase C4 column using gradient elution with acetonitrile and water (Figure 4a). As a result, three fluorescent fractions (designated fractions 1−3 in Figure 4a) were obtained, which were analyzed by a nanoliquid reverse phase C18 chromatography (LC) system equipped with an LED-induced fluorescent detector (FLD) connected to a nanoelectrospray-tandem mass spectrometry (nano-ESI-MS^n) system.[33] Among the fluorescent peaks observed by LC-FLD chromatograms (Figure 4b), only peak 1, which was eluted in fraction 3 of the reverse-phase C4 chromatogram, gave an assignable MS signal that exhibits a characteristic spectrum of the glycopeptide following MS/MS analyses (Figure 4c,d). As other peaks did not show any signal characteristic of glycopeptide (Figure S4 of the Supporting Information), they were not analyzed further. As shown in Figure 4c, peak 1 generated a doubly charged ion at m/z 1435.3, which corresponds to a doubly protonated ion of glycopeptide **6** (Figure 5a) (calcd for $[M + 2H]^+$ m/z 1435.1). This glycopeptide was assigned as squaramide formed between the amino group of K1424 and probe **1** and by the subsequent tryptic cleavages at the peptide bonds between K1423 and K1424 and between R1426 and K1427 (Figure 5b). The structure was confirmed by MS/MS analysis of the precursor ion (m/z 1435.3), which showed a fragment ion resulting from the loss of $Man_9GlcNAc$ at m/z 1207.4 (Figure 4d).

DISCUSSION

Because of its ability to discriminate the folding state of glycoproteins, UGGT functions as a folding sensor in the ER. It preferentially accepts incompletely folded glycoproteins having nonglucosylated high-mannose-type glycans. Whereas this specificity is pertinent to its function in the glycoprotein folding cycle, the molecular basis of its unique property has been poorly understood. As an exposed hydrophobic patch is a hallmark of folding defective proteins, it would be most logical to speculate that hydrophobic interaction is responsible for its substrate recognition.[19] The fact that the enzyme itself has a surface-exposed hydrophobicity reinforces the speculation.[34] However, no extensive effort has been made to narrow the region that plays a role in substrate recognition.

In this study, we made at attempt to search the aglycon recognition site(s) of UGGT by means of affinity labeling. For this purpose, a tripartite oligosaccharide probe **1** was synthesized, which is composed of a high-mannose-type oligosaccharide, a fluorophoric aglycon, and a cross-linking handle equipped with a squaric acid amide ester. Although squarate chemistry has been extensively employed for conjugation of biomolecules to proteins,[23−25] its application to site-selective labeling of proteins is nearly without precedent.

Indeed, probe **1** was able to label the enzyme in the presence of UDP. By contrast, the labeling was barely detectable in the absence of UDP, excluding the possibility that the labeling was caused by nonspecific binding of **1** to the enzyme. Somewhat unexpectedly, only extremely faint staining was observed even in the presence of UDP-Glc, a donor substrate of the enzyme. This result may indicate that UDP-Glc, once incorporated into the active site, rapidly delivers Glc residue to **1** to give the glucosylated product, which is immediately released from the enzyme.

To locate the modified site, labeling experiments using the N-terminal half (residues 38−744) as well as the C-terminal half (residues 745−1555) of HUGT1 were conducted (Figure 3). More intense labeling was observed for the latter, indicating that the C-terminal half of the protein was modified predominantly. Although this result reconciles with the recent report by Kato et al., who suggested the role of thioredoxin-like domain 3 (Trx-3) as a folding sensor region,[35] our attempt was continued to identify the modified location, which unexpectedly revealed the involvement of the C-terminal domain in substrate recognition. Namely, LC−MS analysis of a fluorescently active fraction (peak 1) derived from the labeled protein allowed us to assign it as glycan-modified tripeptide **6**. Because the tripeptide sequence

DOI: 10.1021/acs.biochem.5b00785
Biochemistry 2015, 54, 4909−4917

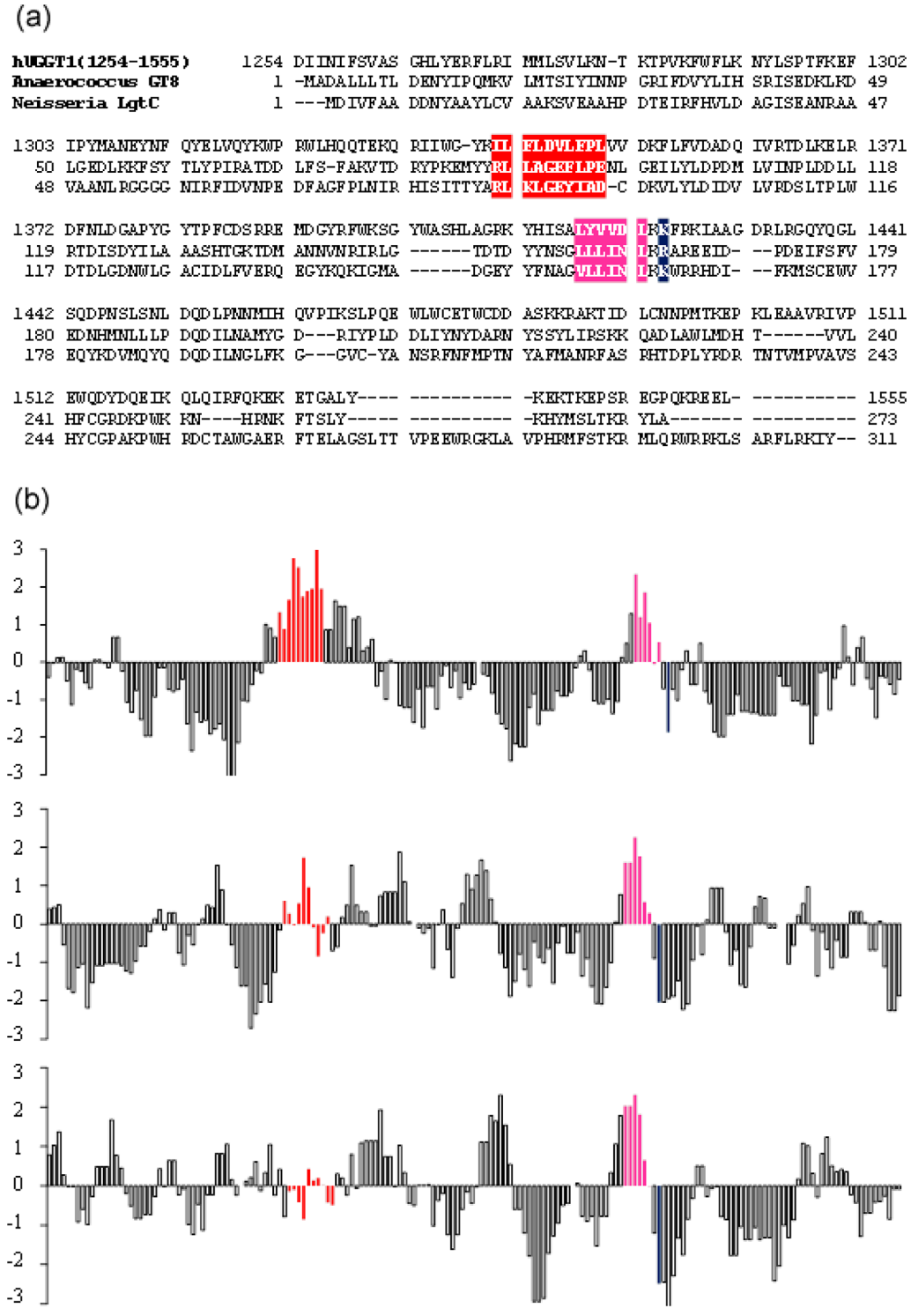


Figure 6. Amino acid sequence profiles of the catalytic region of HUGT1 and glycosyltransferase family 8 (GT8) proteins. (a) Alignment of the HUGT1 and GT8 sequences. (b) Kyte−Doolittle hydrophobicity analysis (seven-residue window). The consensus abbreviations and coloring scheme are as follows. Anaerococcus GT8 indicates the sequences of *A. prevotii* glycosyl transferase family 8 (GenBank entry ACV28465.1); Neiseria LgtC indicates the sequences of *N. meningitidis* LgtC (GenBank entry AAL12839.1). The labeled lysine and the corresponding residues are colored blue; 1340-ILFLDVLFPL-1349 of HUGT1 and the matching sequences are colored red, and 1417-LYVVDL-1422 of HUGT1 and the matching sequences are colored pink.

(KFR) is unique in the sequence of HUGT1, we identified K1424 as the labeled amino acid residue.

In an effort to deduce the location of the aglycon binding site, we searched hydrophobic sequences in the proximity of K1424. For this purpose, analysis based on a Kyte−Doolittle hydropathy plot,[26] which allows prediction of the hydrophobic region of the protein based on its amino acid sequence, was conducted. For comparison, hydropathy plots of other GT-8 family enzymes, *Anaerococcus prevotti* GT8 and *Neisseria meningitides* LgtC, whose tertiary structures have been characterized by X-ray crystallog-

DOI: 10.1021/acs.biochem.5b00785
Biochemistry 2015, 54, 4909−4917

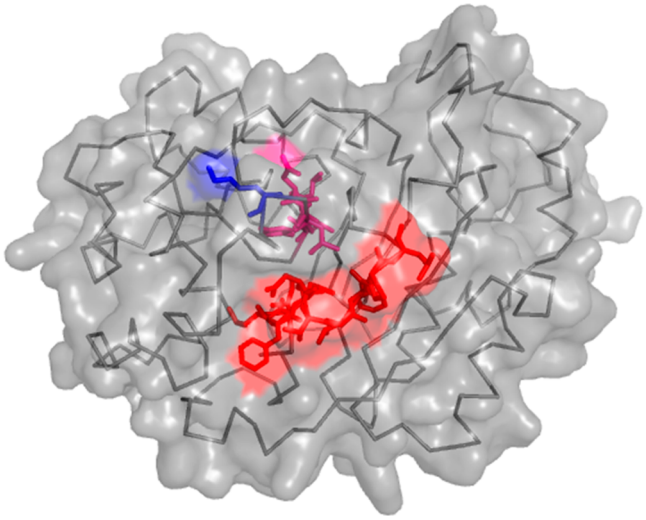


Figure 7. Structural predictions for the catalytic region of HUGT1. Predictive model for HUGT1 residues 1244−1555 (of the 312-residue protein), based on the *A. prevotii* glycosyl transferase family 8 structure (PDB entry 3TZT) and *N. meningitidis* LgtC (PDB entry 1GA8). The labeled lysine is colored blue. Hydrophobic regions 1340-ILFLDV-LFPL-1349 and 1417-LYVVDL-1422 are colored red and pink, respectively.

raphy and whose activities are not affected by the existence of hydrophobic moieties in glycosyl acceptors, were also acquired, as they have high degrees of sequence similarity with the catalytic domain of HUGT1 (Figure 6a). The analysis revealed two hydrophobic regions, 1417-LYVVDL-1422 and 1340-ILFLDV-LFPL-1349 (Figure 6b). The former sequence composed of six amino acids is located only one residue upstream of K1424, while the latter decapeptide lies more distant from the labeled lysine residue. However, hydrophobic segments similar to HUGT1 residues 1417−1422 were also found in *Anaerococccus* GT8 and *Neisseria* LgtC, while both of them are devoid of the additional hydrophobic stretch that corresponds to HUGT1 residues 1340−1349 (Figure 6b). As UGGT is distinct from other GT-8 family enzymes in its ability to recognize the hydrophobic aglycon, the sequence of residues 1340−1349 is likely to be involved in sensing protein folding.

The fact that the K1424 residue was labeled by probe **1** suggests the presence of a folding sensor region in the C-terminal part of UGGT. The conclusion is, however, contrary to the widespread belief that the N-terminal domain serves as the folding sensor. However, it has been postulated that the C-terminal domain of UGGT might be involved not only in the catalytic activity but also in the recognition of the hydrophobic region of substrate protein, while the N-terminal domain facilitates proper folding of the catalytic domain.[8] However, given that UGGT needs to accept various glycoproteins, UGGT might have multiple folding sensor regions consisting of clusters of hydrophobic amino acids. Indeed, our analysis using HUGT1 fragments (Figure 3) provided fluorescent staining of the N-terminal half (residues 38−744) in addition to more intense staining of the C-terminal fragment (residues 745−1555), suggesting the presence of an additional folding sensor domain(s) distant from the catalytic domain. Furthermore, Kato and co-workers recently reported the three-dimensional structure of the third Trx domain (Trx-3) of UGGT derived from *Cheatromium thermophilium*, which revealed the presence of an extensive hydrophobic patch in its flexible C-terminal helix.[34]

As a small molecule probe was used in our study, it is plausible that the hydrophobic region most proximal to the catalytic site was preferentially recognized. Although our analysis identified a unique lysine residue that is likely to be in the proximity of the substrate binding site of the enzyme, it is also possible that the binding pocket is formed by both C-terminal and N-terminal domains, while the squarate moiety was adventitiously oriented toward the amino group of the side chain of K1424. Nevertheless, our results gave an indication that a hydrophobic sequence consisting of 10 amino acids (1340−1349), possibly in combination with another hydrophobic patch(s), serves as one of the folding sensor domains of HUGT1.

Homology modeling analysis of the C-terminal catalytic domain consisting of 312 residues, based on the *A. prevoti* glycosyltransferase family structure (PDB entry 3TZT) and *N. meningitides* LgtC (PDB entry 1GA8), suggested that the canonical folding sensor domain (residues 1340−1349) is exposed on the surface, while the hexapeptide (residues 1417−1422) is buried inside (Figure 7). Intriguingly, the decapeptide (ILFLDVLFPL) is fully conserved among UGGTs derived from *Cheatromium thermophilium*, *Shizosaccharomyces pombe*, *Drosophila melanogaster*, and *Homo sapiens*,[34] reinforcing its importance for the function of these enzymes.

To further substantiate the conclusion, docking simulation studies as well as labeling experiments using HUGT1 mutants will be subjects of future study.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.biochem.5b00785.

Experimental details for the preparation of substrates and affinity labeling assays using compound **1** for HUGT1 (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*RIKEN, Synthetic Cellular Chemistry Laboratory, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan. E-mail: yukito@riken.jp. Fax: (+81)48-462-4680.

Present Addresses

[§]Y.T.: Department of Biotechnology, Ritsumeikan University, Kusatsu, Shiga 525-8577, Japan.

[∥]M.H.: Department of Biological Science and Technology, Tokyo University of Science, Tokyo 125-8585, Japan.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Dr. Yasuhiro Kajihara and Dr. Osamu Kanie for valuable advice and Ms. Akemi Takahashi and Ms. Satoko Shirahata for technical assistance. We also thank the Support Unit for Biomaterial Analysis, RIKEN BSI Research Resources Center.

■ ABBREVIATIONS

CBB, Coomassie brilliant blue; CNX, calnexin; CRT, calreticulin; ESI-MS, electrospray ionization mass spectrometry; ER, endoplasmic reticulum; G-I, glucosidase I; G-II, glucosidase II; Glc, D-glucose; GlcNAc, *N*-acetyl-D-glucosamine; Man, D-

DOI: 10.1021/acs.biochem.5b00785
Biochemistry 2015, 54, 4909−4917

mannose; UDP, uridine diphosphate; UGGT, UDP-glucose:-glycoprotein glucosyltransferase.

REFERENCES

(1) Whitley, P., Nilsson, I., and vonHeijne, G. (1996) A nascent secretory protein may traverse the ribosome endoplasmic reticulum translocase complex as an extended chain. *J. Biol. Chem. 271*, 6241−6244.

(2) Ruiz-Canada, C., Kelleher, D. J., and Gilmore, R. (2009) Cotranslational and posttranslational N-glycosylation of polypeptides by distinct mammalian OST isoforms. *Cell 136*, 272−283.

(3) Bause, E. (1983) Structural requirements of N-glycosylation of proteins - studies with proline peptides as conformational probes. *Biochem. J. 209*, 331−336.

(4) Caramelo, J. J., and Parodi, A. J. (2008) Getting in and out from calnexin/calreticulin cycles. *J. Biol. Chem. 283*, 10221−10225.

(5) D'Alessio, C., Caramelo, J. J., and Parodi, A. J. (2010) UDP-Glc:glycoprotein glucosyltransferase-glucosidase II, the ying-yang of the ER quality control. *Semin. Cell Dev. Biol. 21*, 491−499.

(6) Caramelo, J. J., Castro, O. A., Alonso, L. G., De Prat-Gay, G., and Parodi, A. J. (2003) UDP-Glc:glycoprotein glucosyltransferase recognizes structured and solvent accessible hydrophobic patches in molten globule-like folding intermediates. *Proc. Natl. Acad. Sci. U. S. A. 100*, 86−91.

(7) Parodi, A. J. (2000) Protein glucosylation and its role in protein folding. *Annu. Rev. Biochem. 69*, 69−93.

(8) Guerin, M., and Parodi, A. J. (2003) The UDP-glucose:glycoprotein glucosyltransferase is organized in at least two tightly bound domains from yeast to mammals. *J. Biol. Chem. 278*, 20540−20546.

(9) Taylor, S. C., Thibault, P., Tessier, D. C., Bergeron, J. J. M., and Thomas, D. Y. (2003) Glycopeptide specificity of the secretory protein folding sensor UDP-glucose glycoprotein:glucosyltransferase. *EMBO Rep. 4*, 405−411.

(10) Zapun, A., Petrescu, S. M., Rudd, P. M., Dwek, R. A., Thomas, D. Y., and Bergeron, J. J. M. (1997) Conformation-independent binding of monoglucosylated ribonuclease B to calnexin. *Cell 88*, 29−38.

(11) Keith, N., Parodi, A. J., and Caramelo, J. J. (2005) Glycoprotein tertiary and quaternary structures are monitored by the same quality control mechanism. *J. Biol. Chem. 280*, 18138−18141.

(12) Ritter, C., Quirin, K., Kowarik, M., and Helenius, A. (2005) Minor folding defects trigger local modification of glycoproteins by the ER folding sensor GT. *EMBO J. 24*, 1730−1738.

(13) Totani, K., Ihara, Y., Matsuo, K., Koshino, H., and Ito, Y. (2005) Synthetic substrates for an endoplasmic reticulum protein-folding sensor, UDP-glucose:glycoprotein glucosyltransferase. *Angew. Chem., Int. Ed. 44*, 7950−7954.

(14) Totani, K., Ihara, Y., Tsujimoto, T., Matsuo, I., and Ito, Y. (2009) The recognition motif of the glycoprotein-folding sensor enzyme UDP-Glc:glycoprotein glucosyltransferase. *Biochemistry 48*, 2933−2940.

(15) Ito, Y., Takeda, Y., Seko, A., Izumi, M., and Kajihara, Y. (2015) Functional analysis of endoplasmic reticulum glucosyltransferase-(UGGT): Synthetic chemistry's initiative in glycobiology. *Semin. Cell Dev. Biol. 41*, 90−98.

(16) Takeda, Y., Seko, A., Hachisu, M., Daikoku, S., Izumi, M., Koizumi, A., Fujikawa, K., Kajihara, Y., and Ito, Y. (2014) Both isoforms of human UDP-glucose:glycoprotein glucosyltransferase are enzymatically active. *Glycobiology 24*, 344−350.

(17) Arnold, S. M., Fessler, L. I., Fessler, J. H., and Kaufman, R. J. (2000) Two homologues encoding human UDP-glucose:glycoprotein glucosyltransferase differ in mRNA expression and enzymatic activity. *Biochemistry 39*, 2149−2163.

(18) Arnold, S. M., and Kaufman, R. J. (2003) The noncatalytic portion of human UDP-glucose:glycoprotein glucosyltransferase I confers UDP-glucose binding and transferase function to the catalytic domain. *J. Biol. Chem. 278*, 43320−43328.

(19) Izumi, M., Makimura, Y., Dedola, S., Seko, A., Kanamori, A., Sakono, M., Ito, Y., and Kajihara, Y. (2012) Chemical synthesis of intentionally misfolded homogeneous glycoprotein: a unique approach for the study of glycoprotein quality control. *J. Am. Chem. Soc. 134*, 7238−7241.

(20) Means, G. E., and Feeney, R. E. (1990) Chemical modifications of proteins: history and applications. *Bioconjugate Chem. 1*, 2−12.

(21) Patricelli, M. P., Szardenings, A. K., Liyanage, M., Nomanbhoy, T. K., Wu, M., Weissig, H., Aban, A., Chun, D., Tanner, S., and Kozarich, J. W. (2007) Functional interrogation of the kinome using nucleotide acyl phosphates. *Biochemistry 46*, 350−358.

(22) Cohen, M. S., Zhang, C., Shokat, K. M., and Taunton, J. (2005) Structural bioinformatics-based design of selective, irreversible kinase inhibitors. *Science 308*, 1318−1321.

(23) Tietze, L. F., Arlt, M., Beller, M., Glusenkamp, K. H., Jahde, E., and Rajewsky, M. F. (1991) Anticancer agents 0.15. squaric acid diethyl ester - a new coupling reagent for the formation of drug biopolymer conjugates - synthesis of squaric acid ester amides and diamides. *Chem. Ber. 124*, 1215−1221.

(24) Storer, R. I., Aciro, C., and Jones, L. H. (2011) Squaramides: physical properties, synthesis and applications. *Chem. Soc. Rev. 40*, 2330−2346.

(25) Hou, S. J., Saksena, R., and Kovac, P. (2008) Preparation of glycoconjugates by dialkyl squarate chemistry revisited. *Carbohydr. Res. 343*, 196−210.

(26) Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res. 25*, 4876−4882.

(27) Kyte, J., and Doolittle, R. F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol. 157*, 105−132.

(28) Kelley, L. A., and Sternberg, M. J. E. (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protoc. 4*, 363−371.

(29) Totani, K., Matsuo, I., Ihara, Y., and Ito, Y. (2006) High-mannose-type glycan modifications of dihydrofolate reductase using glycan-methotrexate conjugates. *Bioorg. Med. Chem. 14*, 5220−5229.

(30) Kamitakahara, H., Suzuki, T., Nishigori, N., Suzuki, Y., Kanie, O., and Wong, C. H. (1998) A lysoganglioside poly-L-glutamic acid conjugate as a picomolar inhibitor of influenza hemagglutinin. *Angew. Chem., Int. Ed. 37*, 1524−1528.

(31) Dedola, S., Izumi, M., Makimura, Y., Seko, A., Kanamori, A., Sakono, M., Ito, Y., and Kajihara, Y. (2014) Folding of synthetic homogeneous glycoproteins in the presence of a glycoprotein folding sensor enzyme. *Angew. Chem., Int. Ed. 53*, 2883−2887.

(32) Lairson, L. L., Henrissat, B., Davies, G. J., and Withers, S. G. (2008) Glycosyltransferases: structures, functions, and mechanisms. *Annu. Rev. Biochem. 77*, 521−555.

(33) Daikoku, S., Ono, Y., Ohtake, A., Hasegawa, Y., Fukusaki, E., Suzuki, K., Ito, Y., Goto, S., and Kanie, O. (2011) Fluorescence-monitored zero dead-volume nanoLC-microESI-QIT-TOF MS for analysis of fluorescently tagged glycosphingolipids. *Analyst 136*, 1046−1050.

(34) Sakono, M., Seko, A., Takeda, Y., Hachisu, M., and Ito, Y. (2012) Biophysical properties of UDP-glucose:glycoprotein glucosyltransferase, a folding sensor enzyme in the ER, delineated by synthetic probes. *Biochem. Biophys. Res. Commun. 426*, 504−510.

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

DOI: 10.1021/acs.biochem.5b00785
Biochemistry 2015, 54, 4909−4917