Molecular Recognition

DOI: 10.1002/anie.201309665

Folding of Synthetic Homogeneous Glycoproteins in the Presence of a Glycoprotein Folding Sensor Enzyme**

Simone Dedola, Masayuki Izumi, Yutaka Makimura, Akira Seko, Akiko Kanamori, Masafumi Sakono, Yukishige Ito, and Yasuhiro Kajihara**

Abstract: *UDP-glucose:glycoprotein glucosyltransferase (UGGT) plays a key role in recognizing folded and misfolded glycoproteins in the glycoprotein quality control system of the endoplasmic reticulum. UGGT detects misfolded glycoproteins and re-glucosylates them as a tag for misfolded glycoproteins. A flexible model to reproduce in vitro folding of a glycoprotein in the presence of UGGT in a mixture containing correctly folded, folding intermediates, and misfolded glycoproteins is described. The data demonstrates that UGGT can re-glucosylate all intermediates in the in vitro folding experiments, thus indicating that UGGT inspects not only final folded products, but also the glycoprotein folding intermediates.*

Glycoprotein quality control (GQC)[1] in the endoplasmic reticulum (ER) is a pivotal system for yielding correctly folded glycoproteins instead of misfolded glycoproteins. The glycoprotein folding sensor enzyme, UDP-glucose:glycoprotein glucosyltransferase[2,3] (UGGT), plays a central role in GQC to classify correctly folded and misfolded protein structures (Figure 1). UGGT recognizes hydrophobic patches exposed on the protein surface as an indicator of misfolded structure, and then catalyzes the transfer of a glucose (Glc) residue to the terminus of M9-high-mannose-type oligosaccharides of the misfolded glycoproteins specifically. The resultant misfolded glycoproteins are subsequently passed to the calnexin/calreticulin[4] (CNX/CRT) chaperone system to undergo a refolding process. The glucosylation embodies the critical tag to classify the misfolded glycoproteins. The rigorous substrate specificity of UGGT was confirmed by use of homogeneous, correctly folded and misfolded glycoprotein models, however in those previous reports structurally stabilized glycoprotein models were used.[2,3,5–12] In contrast, glycoprotein folding processes in the ER are a dynamic event and there are potentially many glycoprotein-folding intermediates leading to the formation of the correctly folded (Figure 1E) or misfolded forms (Figure 1F). Almost all glycoproteins possess disulfide bonds and the formation of the correct disulfide bond pair is a critical step in the folding process.

Because of the rate-limiting step of disulfide-bond formation, folding intermediates, which proceed through to the correctly folded glycoprotein, accumulate at a local free-energy minimum (Figure 1C) before the formation of the correct disulfide bond pairs or the final correct protein structure (Figure 1E). The folding process starts from a random coil structure, which can potentially present several types of flexible arrangements as a starting point, and can proceed through two main pathways: a productive folding route (Figure 1A,C,E) which will yield a correctly folded product, and an unproductive folding route (Figure 1A,D,F) which will yield misfolded or aggregated forms.[13,14]

Ideally, to enhance GQC ability, UGGT should not interact with productive folding intermediates that yield correctly folded glycoproteins (Figure 1C), but should interact with those structures (Figure 1D) which are going to produce misfolded or severely aggregated forms (Figure 1F). However, to our knowledge, there is no report describing the ability of UGGT to recognize folding intermediates during the dynamic folding process, but they do recognize stabilized, correctly folded or misfolded substrates. The possibility of examining the above experiments was mainly hampered by two technical drawbacks. Firstly, the difficulty of establishing in vitro folding experiments in the presence of active UGGT, and secondly, the difficulty in designing an experiment including both productive and unproductive folding species in a controlled and reproducible way by use of homogeneous M9-glycosyl polypeptides. Moreover, to gain insight into when UGGT interacts with folding intermediates, the assignment of each folding intermediate included in the productive or unproductive process is essential. However, glycoprotein folding including both productive and unproductive pathways simultaneously, make it difficult to discriminate the interaction of UGGT with folding intermediates. Refolding experiments in vitro usually require guanidine (Gn) or urea. However, because Gn is a denaturizing agent, this also results in the loss of UGGT enzyme activity. Recently, several kinds of homogenous glycoproteins have been synthesized and it was found that the in vitro refolding conditions of relatively large glycoproteins yielded correctly folded glycoproteins

[*] Dr. S. Dedola, Dr. M. Izumi, Dr. Y. Makimura, Prof. Dr. Y. Kajihara
Department of Chemistry, Graduate School of Science
Institution Osaka University
1-1 Machikaneyama-cho, Toyonaka, Osaka, 560-0043 (Japan)
E-mail: kajihara@chem.sci.osaka-u.ac.jp
Dr. S. Dedola, Dr. M. Izumi, Dr. Y. Makimura, Dr. A. Seko, Dr. A. Kanamori, Dr. M. Sakono, Prof. Dr. Y. Ito, Prof. Dr. Y. Kajihara
Institution Japan Science and Technology Agency (JST)
ERATO, Ito Glycotrilogy Project
2-1 Hirosawa, Wako, Saitama, 351-0198 (Japan)
E-mail: yukito@riken.jp
Prof. Dr. Y. Ito
Institution Synthetic cellular Chemistry Laboratory, RIKEN
2-1 Hirosawa, Wako, Saitama, 351-0198 (Japan)

[**] This work was supported by the Japan Science and Technology Agency.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201309665.

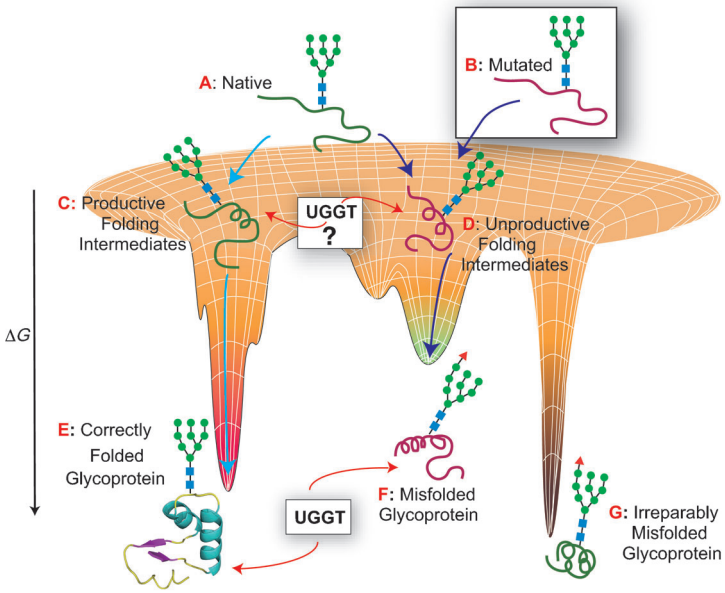

Figure 1. UGGT inspection during the glycoprotein folding process. Native glycoproteins A can follow a productive folding course to yield correctly folded glycoproteins (E), or an unproductive folding course, thus providing misfolded glycoprotein F, which can be re-glucosylated by UGGT for an additional folding attempt and can finally generate irreparably misfolded glycoproteins G, which will eventually be disposed of.

along with misfolded glycoproteins, even in the presence of Gn.[5,15,16] These results clearly indicate that the folding conditions include productive and unproductive folding processes simultaneously, and we could not regulate the folding efficiency to have only the correctly folded glycoproteins. With the aim of overcoming such issues, we found glycosyl crambin (Figure 2a) to be a suitable substrate for our purpose. Glycosyl crambin is a small protein with three disulfide bonds and exhibits suitable productive folding processes which yield a correctly folded glycoprotein as the sole product, even in the absence of Gn or urea. Based on this finding, we set out to prepare both the productive folding substrate **4** and unproductive folding substrate **5**. With respect to **5**, we prepared a mutant wherein two of the six native cysteine residues in the crambin sequence were substituted with serine. The mutation sites were selected so as to have two of the three native disulfide bonds disrupted, with a new pairing pattern which could potentially generate three species with two disulfide bonds each. To investigate whether UGGT could recognize folding intermediates in both the productive and unproductive folding process, we designed a unique UGGT assay which we thought could be a useful tool to mimic the folding in the ER.

Herein, we describe the critical substrate specificity of UGGT towards folding intermediates by using our new strategy which consists of simultaneous productive and unproductive folding in the presence of active UGGT. A mixture of homogeneous glycosyl native crambin and its mutant is used.

For the synthesis of both **4** and its mutant **5**, we employed a segment-coupling strategy with subsequent folding experiments. The cysteine at the 16-position was selected for the junction by native chemical ligation. The target polypeptide chain shown in Figure 2 was divided into two segments 1–15, and 16–46. The Asn-M9-glycan, Fmoc-Asn-(M9-high-mannose type oligosaccharide)-OH, was obtained by isolation from egg yolk and then used for Fmoc solid-phase polypeptide synthesis (SPPS).[17] In terms of the glycosylation position, we selected residue 17 because intentional glycosylation at this loop region should not disrupt the three-dimensional structure of the protein (PDB 1CNR). Segments 1 and 2 were synthesized by *tert*-Boc and Fmoc SPPS, respectively, in moderate yield and native chemical ligation successfully yielded the full glycosyl peptide target. The details of the syntheses are reported in the Supporting Information.

After the synthesis of the glycosylated **4** and **5**, folding experiments were examined. The glycopeptide **4** correctly folded under both conventional redox conditions in the presence of Gn (see Figure S6a in the Supporting Information) and air oxidative folding in the absence redox conditions (see Figure S8a). In contrast, the folding of **5** in the conventional redox system did not proceed to completion (Figure S6b) because of the formation of an unproductive disulfide bond with cysteine

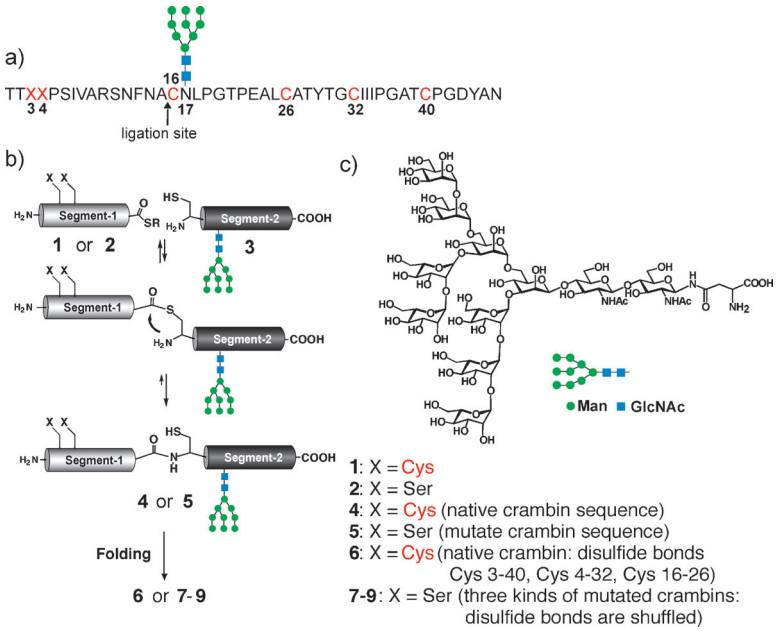

Figure 2. a) Amino-acid sequence of crambin: cysteines are highlighted in red and mutations are indicated by "X". b) General scheme for native chemical ligation (NCL) with the potential number of final products for the correctly folded **6** and mutated folded glycosyl crambins **7**, **8**, and **9** (three possible combinations for the four cysteines in the mutant **5**). c) Tri-antennary high-mannose oligosaccharide.

used in the redox buffer. Therefore we selected air oxidative folding in the absence redox conditions for **5** (Figure S8b). To observe folding of **4** and **5** in the presence of UGGT, we had to confirm that the folding of those species was possible under conventional UGGT assay conditions. We found suitable conditions to be a Tris-HCl 100 mM buffer at pH 7.5, which is a standard for a UGGT assay.[18] As shown in Figure 3, **4** indeed folded into a single product (**6**) at pH 7.5 (Figure 3a), and its mutant yielded the three expected species **7**, **8**, and **9** (Figure 3b).

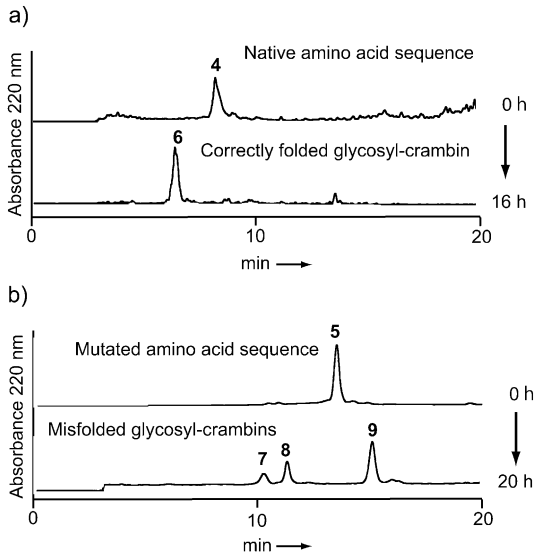

Figure 3. HPLC profiles for air oxidative folding of **4** and **5** in UGGT buffer at pH 7.5 (in the absence of UGGT): a) Folding of **4** at pH 7.5, folding time was 16 h. b) Folding of **5** at pH 7.5, folding time was 20 h.

To characterize the products **6** and **7**–**9** derived from **4** and **5**, respectively, we employed electrospray ionization (ESI)-mass analysis, ^{1}H-^{1}H TOCSY spectroscopy, and circular dichroism analysis. In the case of **4**, ESI/MS showed a suitable mass pattern and a loss of six daltons compared with the mass of the starting M9-glycosyl polypeptide, thus indicating the formation of three disulfide bonds (see Figure S7b). The ^{1}H-^{1}H TOCSY spectrum of **6** exhibited the same chemical shift pattern as that of a previously synthesized and correctly folded glycosyl crambin having a complex type oligosaccharide[15] (see Figure S21). This TOCSY spectrum shows that glycosyl crambin possessing a native sequence employs a suitable three-dimensional protein conformation. In terms of all three misfolded products (**7**–**9**) derived from **5**, a decrease in mass of four daltons, compared with that of starting M9-glycosyl polypeptide mutant (see Figure S9b), was observed. This data means that the three products are structural isomers with each one having a different disulfide bond pairing. CD spectra of those crambins also showed suitable spectral patterns. Correctly folded **6** exhibited both helices and sheet patterns at the $\lambda = 209$ and 220 nm, respectively (see Figure S11a). In contrast, the shape of the CD spectra of **7**–**9** more closely resemble that of a random coil structure (Figure S11b) compared to that of the correctly folded **6**, thus confirming the unstructured nature of the folded mutants and their suitability to mimic misfolded glycoproteins in our assay.

It is known that UGGT recognizes the hydrophobic patches on protein surfaces, therefore we examined an ANS binding assay,[19] which is considered to be the gauge of hydrophobicity of a protein surface. Such hydrophobicity can be estimated by the fluorescence intensity resulting from an ANS–protein complex (see Figure S12c). The ANS assay indicated that the three misfolded and one correctly folded crambin exhibited hydrophobic patches on the protein surface, and that the hydrophobicity for each was comparable and smaller than that of the misfolded glycosyl interleukin-8 dimer previously synthesized.[5]

To investigate how UGGT recognizes those synthetic products, UGGT assays ford each glycoprotein were examined. The enzymes used in those experiments were human UGGT (Figure 4a,b) and *A. oryzae* UGGT (Figure 4c,d) prepared from *E. coli* expression.

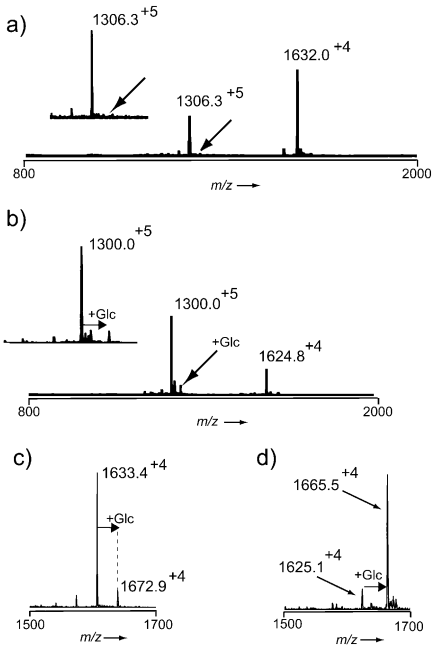

Figure 4. ESI-MS spectrum for the human UGGT reaction product of **6** (a) and **9** (b). ESI-MS spectrum for the *A. oryzae* UGGT reaction product of **6** (c) and **9** (d). The arrow indicates the transferase product. For a comparison between the reaction at the start and after 24 h see the ESI-MS spectra in Figures S13 (human UGGT) and S14 (*A. oryzae* UGGT).

Because the product of the UGGT assay, G1M9-glycoproteins, cannot be separated from the corresponding M9-glycoprotein substrates on the basis of an HPLC profile, LC-ESI mass spectrometry was used to estimate the formation of G1M9-glycoproteins by the peak intensity of each mass signal (Figure 4). We employed the conventional UGGT assay solution previously reported.[5,18] In short, a Tris-HCl buffer solution (100 mM, pH 7.5) containing glycoprotein (7.5 μM), UDP-Glc (0.5 mM), $CaCl_2$ (5 mM), and UGGT were incubated at 37 °C. After direct injection of the reaction mixture

into the LC/MS, G1M9-glycoproteins were measured by the intensity of signals for the mass peaks. As expected, UGGT was found to transfer a Glc residue to the three misfolded M9-glycosyl crambins (Figure 4b,d; products are highlighted by arrows), while the correctly folded M9-glycosyl crambin was not significantly glucosylated (Figure 4a,c). The misfolded M9-glycosyl crambins accepted 20% and 80% of the Glc residue in the case of human and *A. oryzae* UGGT, respectively, whereas the correctly folded M9-glycosyl crambin accepted less than 2% and 8% of Glc in the case of human and *A. oryzae* UGGT, respectively, as determined by comparison to that of the parent M9-glycosyl crambin.

These results enabled us to examine in the following experiments into whether UGGT can recognize folding intermediates. As shown in Figure 3a, **4** can fold without Gn and a thiol disulfide redox reagent. Therefore we could design an assay where both polypeptides of **4** and **5** were combined to mimic an ER glycoprotein folding process. In fact, the active UGGT can potentially interact with all productive and unproductive folding intermediates (Figure 5). To maintain UGGT activity, the assay employed air oxidative folding conditions in the absence of Gn or urea as shown in Figure 5. The degree of glucosylation was observed by HPLC during the folding process and were analyzed by LC-ESI MS (Figure 5 and see Tables S1 and S2). The vertical axis indicates the yield of the glucosyl transfer compared to that of the parent polypeptide M9-glycosyl crambin or its mutant, and it was monitored for 12 hours from the start of folding. The enzymes used in those experiments were human UGGT (Figure 5a) and *A. oryzae* UGGT (Figure 5b) prepared from *E. coli* expression. Because the specific activity of the human UGGT was lower than that of *A. oryzae*, we needed to increase the amount of human UGGT enzyme, and thus added a significant amount of enzyme stock solution in the reaction mixture to give rather basic (pH 8.5) conditions, which led to a faster folding velocity. As shown in Figure 5a,b, both human and *A. oryzae* UGGT exhibited suitable glucosyl transfer toward all folding intermediates, that is, the three misfolded and the correctly folded M9-glycosyl crambins. The multiple peaks observed in HPLC profile indicate the folding intermediates in stepwise formation of disulfide bonds.

In the HPLC profiles in Figure 5a,b, the resulting species present during the folding experiment were identified as follows: folded native **6**, a productive folding intermediate with two disulfide bonds (peak a) and one disulfide bond (peak b and c), unfolded **4** and **5**, an unproductive folding intermediate with one disulfide bond (peak d, e, and f), folded and misfolded mutants **7**, **8**, and **9**. The peaks indicated in red in Figure 5a,b are starting material and product of the productive folding pathway as confirmed by the reference experiments shown in Figure 3a.

By comparing the two assays shown in Figures 4 and 5, the correctly folded **6** (Figure 4a) was a very poor substrate for UGGT, thus showing no Glc transfer with human UGGT (close to noise level, Figure 4a) and 8% (Figure 4c) transfer with *A. oryzae* UGGT. The same species **6** was glucosylated at over 6% and at 40% using the human UGGT (Figure 5a, Table S1) and the *A. oryzae* UGGT (Figure 5b, Table S2), respectively, during the folding process.

Although human UGGT did not transfer a large amount of the Glc residue to the correctly folded glycosyl crambin, the glucosyl rate was clearly increased from less than 1% to 6%. Therefore we can conclude that the presence of the Glc on the correctly folded **6** is the result of the UGGT activity on its folding intermediates which finally attained the folded structure with the addition of a Glc. Since we could detect glucosyl transfer for all the intermediates (Figure 5), we conclude that UGGT can detect the folding intermediates in which the protein structure is insufficiently folded or resembles a misfolded form.

Some protein folding is complete within a few milliseconds, but some folding intermediates may accumulate depending upon the protein sequence and the conditions used. In terms of folding of the glycoprotein with cysteines, the formation of disulfide bonds during fold-

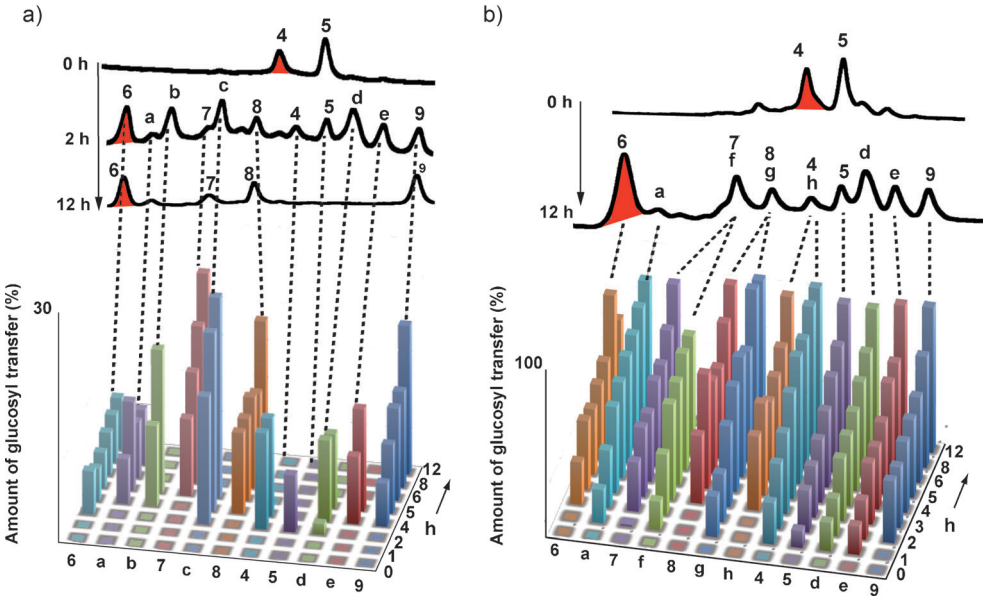

Figure 5. Bar chart for the quantity of Glc transferred to the intermediates during the folding of **4** and **5** in the presence of human UGGT (a) and *A. oryzae* UGGT (b; values are reported in Tables S1 and 2, respectively). The disappearance of the percentage bar in (a) is not related to a decrease of Glc transferred to the corresponding species, but rather to the specific intermediate attaining a final folded structure (i.e. peak b converging to **6**, or peak c converging to **7**). The HPLC peak indicated red is the native glycosyl crambin and the correctly folded glycosyl crambin. This pathway leads to productive folding. Because the folding velocity was different for the data in Figures 5a and 5b, we observed new intermediates such as f–h.

ing is one of the major steps and seems to be the rate-limiting step which results in the accumulation of folding intermediates at local free-energy minima, and cannot be overcome to move on to the next disulfide bond-formation step.[20,21] Therefore, in the ER GQC system, these folding intermediates seem to be recognized by UGGT, glucosylated, and then passed to CNX/CRT chaperones, which recognize G1M9-glycan misfolded glycoproteins and facilitate a refolding process.

The ability of glycosyl crambin to fold without urea or Gn allowed us to investigate the discrimination ability of UGGT at any time point during the glycoprotein folding process. The choice of crambin as a model was also made in consideration of its well-known hydrophobic properties. As reported by a number of previous publications,[5–12] UGGT recognizes the exposed hydrophobic patches present in the misfolded glycoproteins. Although an ANS assay indicated slight hydrophobicity on both the correctly folded and misfolded glycosyl crambins, folding intermediates might expose hydrophobic amino acids on the protein surface and therefore UGGT might recognize it. Because UGGT does not seem to recognize consensus sequence of N-glycosylation,[22] we did not insert it. We thought the introduction of a mutation, in addition to the replacement of cysteine with serine, in **5** might yield inefficient folding process. Although the mutation in the amino-acid sequence of hydrophobic proteins frequently yields disordered protein aggregation, the rational mutation that we selected for crambin enabled us to yield three different misfolded glycoproteins. By performing the folding in the presence of UGGT, all the states the model glycoprotein goes through, from the unfolded to the folded/misfolded one, are available to the UGGT for recognition. It should also be mentioned that folding of glycoproteins, containing cysteines, in the ER is assisted by protein disulfide isomerase (PDI), which was not included in our folding assay. Future studies will include several enzymes such as PDI and chaperones to mimic in vitro ER-like folding experiments.

Our unique experiments demonstrated that the specificity of UGGT is therefore not restricted to the final misfolded glycoproteins but also includes all the potential folding intermediates accumulated during the folding process. The design of this rational folding experiment as a mimic of the folding in the ER will be useful for studying other recognition events in dynamic glycoprotein folding.

Received: November 6, 2013
Revised: December 20, 2013
Published online: February 5, 2014

Keywords: biological activity · glycoproteins · mass spectrometry · molecular recognition · protein folding

[1] L. Ellgaard, A. Helenius, *Nat. Rev. Mol. Cell Biol.* **2003**, *4*, 181–191.
[2] M. C. Sousa, M. A. Ferrero-Garcia, A. J. Parodi, *Biochemistry* **1992**, *31*, 97–105.
[3] S. E. Trombetta, M. Bosch, A. J. Parodi, *Biochemistry* **1989**, *28*, 8108–8116.
[4] F. E. Ware, A. Vassilakos, P. A. Peterson, M. R. Jackson, M. A. Lehrman, D. B. Williams, *J. Biol. Chem.* **1995**, *270*, 4697–4704.
[5] M. Izumi, Y. Makimura, S. Dedola, A. Seko, A. Kanamori, M. Sakono, Y. Ito, Y. Kajihara, *J. Am. Chem. Soc.* **2012**, *134*, 7238–7241.
[6] N. Keith, A. J. Parodi, J. J. Caramelo, *J. Biol. Chem.* **2005**, *280*, 18138–18141.
[7] C. Ritter, A. Helenius, *Nat. Struct. Biol.* **2000**, *7*, 278–280.
[8] M. Sakono, A. Seko, Y. Takeda, M. Hachisu, Y. Ito, *Biochem. Biophys. Res. Commun.* **2012**, *426*, 504–510.
[9] S. C. Taylor, P. Thibault, D. C. Tessier, J. J. M. Bergeron, D. Y. Thomas, *EMBO Rep.* **2003**, *4*, 405–411.
[10] D. C. Tessier, D. Dignard, A. Zapun, A. Radominska-Pandya, A. J. Parodi, J. J. Bergeron, D. Y. Thomas, *Glycobiology* **2000**, *10*, 403–412.
[11] K. Totani, Y. Ihara, T. Tsujimoto, I. Matsuo, Y. Ito, *Biochemistry* **2009**, *48*, 2933–2940.
[12] K. Totani, Y. Ihara, I. Matsuo, H. Koshino, Y. Ito, *Angew. Chem.* **2005**, *117*, 8164–8168; *Angew. Chem. Int. Ed.* **2005**, *44*, 7950–7954.
[13] J. L. Arolas, F. X. Aviles, J.-Y. Chang, S. Ventura, *Trends Biochem. Sci.* **2006**, *31*, 292–301.
[14] E. Welker, M. Narayan, W. J. Wedemeyer, H. A. Scheraga, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 2312–2316.
[15] Y. Kajihara, Y. Tanabe, S. Sasaoka, R. Okamoto, *Chem. Eur. J.* **2012**, *18*, 5944–5953.
[16] M. Murakami, R. Okamoto, M. Izumi, Y. Kajihara, *Angew. Chem.* **2012**, *124*, 3627–3632; *Angew. Chem. Int. Ed.* **2012**, *51*, 3567–3572.
[17] Y. Makimura, T. Kiuchi, M. Izumi, S. Dedola, Y. Ito, Y. Kajihara, *Carbohydr. Res.* **2012**, *364*, 41–48.
[18] S. M. Arnold, L. I. Fessler, J. H. Fessler, R. J. Kaufman, *Biochemistry* **2000**, *39*, 2149–2163.
[19] G. V. Semisotnov, N. A. Rodionova, O. I. Razgulyaev, V. N. Uversky, A. F. Gripas, R. I. Gilmanshin, *Biopolymers* **1991**, *31*, 119–128.
[20] W. J. Lees, *ChemBioChem* **2012**, *13*, 1725–1727.
[21] N. Metanis, D. Hilvert, *Angew. Chem.* **2012**, *124*, 5683–5686; *Angew. Chem. Int. Ed.* **2012**, *51*, 5585–5588.
[22] J. J. Caramelo, O. A. Castro, L. G. Alonso, G. de Prat-Gay, A. J. Parodi, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 86–91.

