VIP Protein Folding Very Important Paper

International Edition: DOI: 10.1002/anie.201511491
German Edition: DOI: 10.1002/ange.201511491

Synthesis of Glc_1Man_9-Glycoprotein Probes by a Misfolding/Enzymatic Glucosylation/Misfolding Sequence

Masayuki Izumi, Yukiho Oka, Ryo Okamoto, Akira Seko, Yoichi Takeda, Yukishige Ito, and Yasuhiro Kajihara**

Abstract: *Glycoproteins in non-native conformations are often toxic to cells and may cause diseases, thus the quality control (QC) system eliminates these unwanted species. Lectin chaperone calreticulin and glucosidase II, both of which recognize the Glc_1Man_9 oligosaccharide on glycoproteins, are important components of the glycoprotein QC system. Reported herein is the preparation of Glc_1Man_9-glycoproteins in both native and non-native conformations by using the following sequence: misfolding of chemically synthesized Man_9-glycoprotein, enzymatic glucosylation, and another misfolding step. By using synthetic glycoprotein probes, calreticulin was found to bind preferentially to a hydrophobic non-native glycoprotein whereas glucosidase II activity was not affected by glycoprotein conformation. The results demonstrate the ability of chemical synthesis to deliver homogeneous glycoproteins in several non-native conformations for probing the glycoprotein QC system.*

Protein N-glycosylation occurs in the lumen of the endoplasmic reticulum (ER). Newly synthesized glycoproteins fold into their native structures with the aid of molecular chaperones and folding factors which are resident in the ER. Lectin chaperone calnexin (CNX)/calreticulin (CRT) recognizes the $Glc_1Man_9GlcNAc_2$ (G1M9) N-glycan on a glycoprotein. CNX/CRT is able to assist glycoprotein folding by preventing premature aggregation and recruiting the thiol-oxidoreductase ERp57, which catalyzes disulfide bond formation and rearrangement.[1] Glucosidase II (G-II) then removes a glucose residue, thus converting the N-glycan structure into $Man_9GlcNAc_2$ (M9) so that the glycoprotein can dissociate from the chaperone. At this point, the folding sensor enzyme UDP-glucose:glycoprotein glucosyltransferase (UGGT) inspects the folding status of the glycoproteins: glycoproteins in the native conformation escape the inspection and transit through the secretory pathway. However, glycoproteins in non-native conformations are trapped by UGGT, which then converts the M9 N-glycan on the non-native glycoprotein back into G1M9 so that glycoprotein can reassociate with CNX/CRT for an attempt at refolding.[2]

The process, consisting of CNX/CRT, G-II, and UGGT, is called the CNX/CRT cycle (Figure 1) and it has a major role in the glycoprotein quality control (QC) process,[3] the process that is vital for maintaining protein homeostasis and minimizing pathogenic occurrence.[4] Thus understanding of the detailed mechanism of this cycle is essential for the development of therapeutics for diseases associated with this cycle.[5] In the CNX/CRT cycle, the folding sensor enzyme UGGT is the key component that distinguishes native and non-native conformers of glycoproteins.[6] In contrast, the roles of CNX/CRT chaperones and G-II in the context of discriminating between non-native and native conformers are less clear. Lectin activity of CNX/CRT binding to G1M9 oligosacchar-

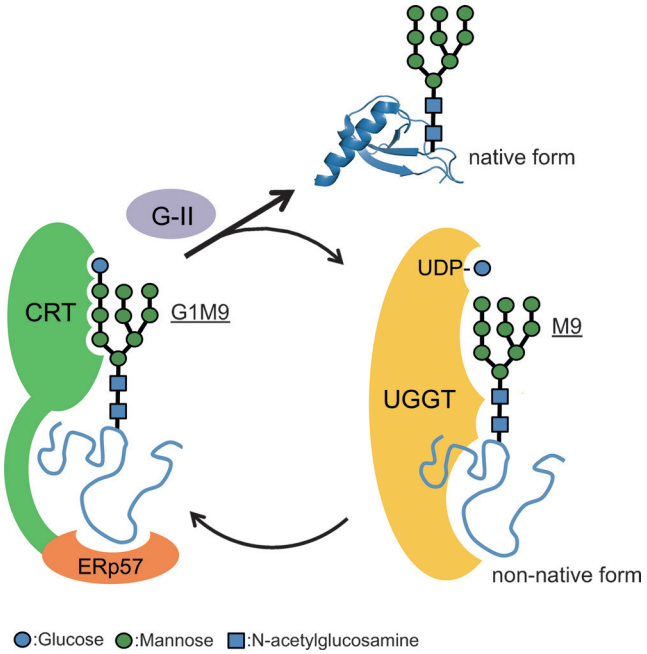


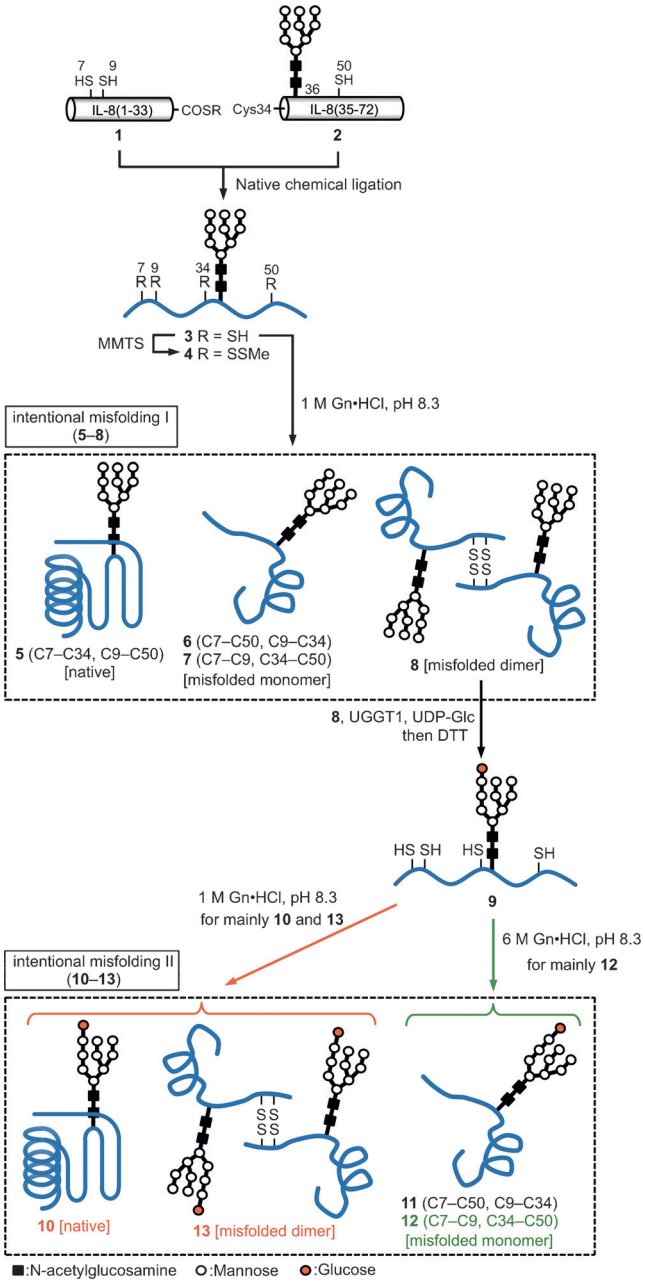
Figure 1. Calnexin/calreticulin cycle for glycoprotein quality control in the endoplasmic reticulum. Lectin chaperone calreticulin (CRT) binds to a glycoprotein bearing the G1M9 N-glycan for refolding with the aid of protein disulfide isomerase ERp57. Glucosidase II (G-II) hydrolyzes a glucose residue from G1M9 N-glycan, thus converting it into M9 N-glycan. UDP-glucose:glycoprotein glucosyltransferase (UGGT) is a folding sensor enzyme which transfers a glucose residue only to a non-native form of glycoprotein bearing an M9 N-glycan.

[*] Dr. M. Izumi, Y. Oka, Dr. R. Okamoto, Prof. Dr. Y. Kajihara
Department of Chemistry, Graduate School of Science
Osaka University
1-1 Machikaneyama, Toyonaka, Osaka 560-0043 (Japan)
E-mail: kajihara@chem.sci.osaka-u.ac.jp
Dr. A. Seko, Dr. Y. Takeda, Prof. Dr. Y. Ito, Prof. Dr. Y. Kajihara
ERATO Ito glycotrilogy project Japan Science and Technology Agency (JST), 2-1 Hirosawa, Wako, Saitama 351-0198 (Japan)
E-mail: yukito@riken.jp
Prof. Dr. Y. Ito
Synthetic Cellular Chemistry Laboratory, RIKEN
2-1 Hirosawa, Wako, Saitama 351-0198 (Japan)
Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under http://dx.doi.org/10.1002/anie.201511491.

ide is well established, although there are different views on the existence of binding sites on non-native polypeptides.[7] Substrate specificity of G-II has been studied in detail, mainly by using an oligosaccharide as the substrate,[8] but the effect of the folding status of a substrate glycoprotein on hydrolysis activity has merely been mentioned. G1M9 glycoprotein probes in different folding states are required to answer these questions. However, these are extremely difficult to obtain from cells as a single glycoform because most of the glycoproteins bearing G1M9 N-glycans are thought to exist only transiently as biosynthetic intermediates. To overcome this difficulty, Ito and co-workers prepared glycoprotein mimics using chemically synthesized G1M9 oligosaccharides and used them for the study of glycan recognition by CRT and G-II.[9] After obtaining GlcNAc-ribonuclease (RNase) from a natural source using endoglycosidase, Wang et al. prepared G1M9-RNase by enzymatic transfer of a chemically synthesized oligosaccharide to GlcNAc-RNase. This approach allowed them to study G1M9-RNase binding to CRT.[10] These elegant studies addressed the effects of oligosaccharide structures, but the effects of protein conformations ought to be elucidated.

We have been employing chemical glycoprotein synthesis to study the QC system of glycoproteins[11] because chemical synthesis is an effective approach to obtain a glycoprotein in a single glycoform.[12] In addition, we can prepare proteins in different conformations, such as in native or non-native (misfolded, unfolded, folding intermediate) forms, using chemical methods. Herein we report the syntheses of G1M9-interleukin 8 (IL-8) in both the native and misfolded forms, including monomers and a dimer wherein the disulfide bonds are scrambled. Since isolating G1M9 oligosaccharides from a natural source in sufficient amounts for use in solid-phase glycopeptide synthesis is cumbersome and time consuming, we used UGGT to convert non-native M9-IL-8, which was prepared by intentional misfolding, into G1M9-IL-8 with high conversion. Subsequently, we carried out another intentional misfolding step to obtain misfolded forms of G1M9-IL-8. Through this enzymatic glucosylation and iterative misfolding approach, we successfully obtained enough G1M9-glycoprotein probes for use in the analysis of the abilities of CRT and G-II in distinguishing conformational differences in glycoproteins.

We first examined enzymatic glucosylation of two M9-IL-8 derivatives, the monomer **4** and dimer **8**, by UGGT1 (Scheme 1). In our previous study, we found that the misfolded M9-IL-8 dimer **8**, bearing two M9 N-glycans, is a very good substrate of UGGT1.[11a] However, the yield of **8** under air oxidation conditions is relatively low because undesired monomers (**5**–**7**) are also formed. Therefore we examined the **4**, wherein the cysteine side-chain thiol groups were protected with methylthio groups, as a substrate of UGGT1. The M9-IL-8 full-length polypeptide **3**[11a] was prepared by connecting the N-terminal IL-8(1-33) peptide thioester **1** and C-terminal glycopeptide IL-8(34-72, N36-(M9)) **2** by native-chemical ligation.[13] Treatment of **3** with *S*-methyl methanethiosulfonate (MMTS) gave **4** in 80% yield (see Figure S1 in the Supporting Information). Then, enzymatic glucosylation of **4** with recombinant human UGGT1 was examined in 50 mM Tris-HCl (pH 7.5), 5 mM $CaCl_2$, and 0.5 mM UDP-glucose at 37 °C. Unfortunately, **4** was not quite stable under the enzymatic reaction conditions, and conversion of **4** into the native form **5** was observed during UGGT1 reaction. Since **5** was not a substrate of UGGT1, we could not achieve good conversion into **4** (see Figure S2 in the Supporting Information).



Scheme 1. Syntheses of G1M9-IL-8 probes in native and misfolded forms by iterative intentional misfolding process. [Intentional misfolding I] Air oxidation of **3** in 1 M Gn·HCl at pH 8.3 gave **5**, **6**, and **7** (SS bonds are indicated in brackets), as well as **8** (interchain SS bond: C(7,9)–C(7,9), intrachain SS bond: C34–C50; blue solid line represents IL-8 polypeptide chain). Enzymatic glucosylation of **8** by recombinant UGGT1 gave **9** after reduction with DTT. [Intentional misfolding II] Air oxidation of **9** in 1 M Gn·HCl at pH 8.3 (red arrow) gave **10**, **11**, **12**, and **13**. Air oxidation of **9** in 6 M Gn·HCl at pH 8.3 (green arrow) **12** as the main product (see Figure 2).

We thus decided to use **8** for the syntheses of G1M9-IL-8 derivatives. The dimer **8** was obtained by intentional misfolding of **3** in air-bubbled 1M guanidine hydrochloride (Gn·HCl) and 0.1M Tris-acetate (pH 8.3), along with monomers in native (**5**) and misfolded (shuffled disulfide bond) forms (**6** and **7**), as previously reported (Scheme 1: intentional misfolding I; see Figure S3 in the Supporting Information). We omitted redox reagents such as cysteine-cystine to suppress rearrangement of kinetically formed disulfide bonds, in the misfolded species, to native disulfide bonds. The dimer **8** was isolated in 19% yield by HPLC. Glucose transfer to **8** was then examined under the same reaction conditions as **4**. Unlike **4**, conversion of **8** into **5** was not observed during UGGT1 reaction. After treating an aliquot of the UGGT1 reaction solution with DL-dithiothreitol (DTT) to convert the reaction product into the monomer **9**, conversion was estimated to be about 89% after 2 hours, based on the intensities of the MS spectrum. This yield did not increase after a prolonged time (see Figure S4 in the Supporting Information). The G1M9-IL-8 polypeptide **9** was isolated in 35% yield after treatment with DTT.

To obtain G1M9-IL-8 derivatives in both native and misfolded forms, we next examined another intentional misfolding step using **9** under two different conditions (Scheme 1:intentional misfolding II). First, we employed the same misfolding conditions described for **3**. Although the oligosaccharide structure was slightly different, air oxidation of **9** in 1M Gn·HCl gave essentially the same result observed for **3**. The reaction yielded the native form **10**, two misfolded monomers (**11** and **12**), and the misfolded dimer **13** (Figure 2a). Since **11** and **12** were only minor products under these conditions, we sought reaction conditions which gave misfolded monomers preferentially. When we increased the concentration of Gn·HCl to 6M, **12** was formed as a major product (Figure 2b). We speculated that high concentrations of Gn·HCl might have stabilized the monomer having one non-native disulfide bond between C34 and C50, thus preventing the formation of dimer **13**. Using these two intentional misfolding conditions, we were able to obtain enough of the native **10**, misfolded monomer **12**, and misfolded dimer **13** for further biochemical studies.

With **10**, **12**, and **13** in hand, we examined the effects of the conformation of the protein moiety on the hydrolysis activity of G-II and ligand-binding activity of the lectin chaperone CRT. The derivatives **10**, **12**, and **13** were incubated with recombinant G-II. Glucose trimming yields by G-II were estimated from MS intensities (Figure 3; black bars). No significant difference in the hydrolysis yield was observed for **10**, **12**, and **13**, thus suggesting that G-II hydrolyzes any glycoprotein bearing the G1M9 N-glycan regardless of its protein conformation. We then examined the binding of **10**, **12**, and **13** with CRT by using inhibition activity of CRT against G-II hydrolysis (Figure 3).[9a] We used recombinant human CRT and mutant CRT K111H, both expressed in *E. coli* (see the Supporting Information). K111 of CRT is known to be involved in the binding of the G1M9 oligosaccharide to CRT, therefore, the mutant K111H has diminished lectin activity.[14] Hydrolyses of both **10** and **12** were inhibited by CRT. However, as hydrolyses of both were not inhibited by

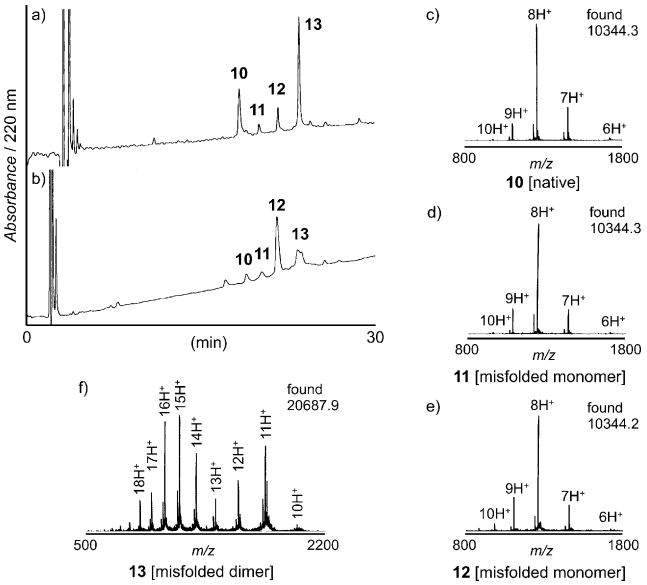


Figure 2. HPLC profiles of the misfolding II step for **9** in a) Air-bubbled 1M Gn·HCl, 0.1M Tris-AcOH, pH 8.3. b) Air-bubbled 6M Gn·HCl, 0.1M Tris-AcOH, pH 8.3. ESI-MS spectra of G1M9-IL-8 probes. c) **10**:$[M+H]^+$ found 10344.3, calcd 10343.4. d) **11**:$[M+H]^+$ found 10344.3, calcd 10343.4. e) **12**:$[M+H]^+$ found 10344.2, calcd 10343.4. f) **13**:$[M+H]^+$ found 20687.9, calcd 20688.0.

CRT K111H, thus suggesting that binding of **10** and **12** to CRT were mediated by the G1M9 N-glycan. In the case of **13**, both CRT and CRT K111H inhibited the hydrolysis by G-II to the similar degree. CRT is reported to be capable of suppressing aggregation of nonglycosylated proteins,[15] and this suppression is reported to be mediated by a hydrophobic peptide binding site.[16] We speculated that since only **8** possesses a hydrophobic patch on the protein surface, as suggested by an 1-anilino-8-naphthalenesulfonate (ANS) binding assay,[11a] the misfolded **13** also possesses a hydrophobic patch on the protein surface and binds not only to the lectin site but also to the polypeptide binding site of CRT.

By observing the in vitro assays using our precisely synthesized native and misfolded G1M9-IL-8 glycoprotein probes, we are able to suggest the role of CRT and G-II in glycoprotein QC process as follows: Wang et al. reported that both native and completely denatured G1M9-RNases bind to CRT with similar affinities,[10] and our results of **10** and **12** also

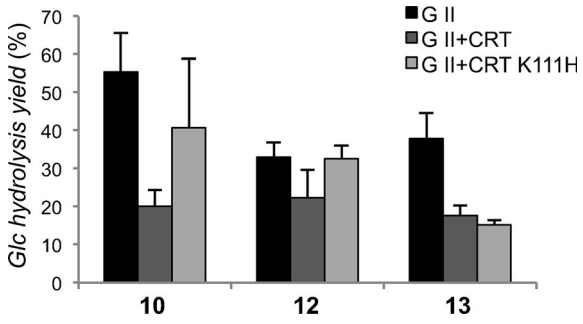


Figure 3. Yields for the Glc hydrolysis of **10**, **12**, and **13** by immobilized G-II after 20 min in either the absence or presence of either CRT or the CRT K111H mutant.

suggest that CRT does not distinguish between native and non-native glycoproteins as UGGT does. However, binding of **13** to lectin-deficient CRT (CRT K111H) suggests the unique mechanism of CRT. CRT is proposed to interact with glycoproteins not only through the lectin site but also through the polypeptide binding site, a phenomenon known as the dual-binding model.[1b] We speculated that the hydrophobic **13** binds to CRT K111H through the polypeptide binding site, thus causing hydrolysis inhibition. Our findings using misfolded monomers and dimers suggest that CRT has binding preference to hydrophobic non-native glycoproteins, although it does not distinguish between native and non-native glycoproteins as strictly as UGGT does. Because both the native and non-native forms of glycoproteins bearing the G1M9 N-glycan bind to CRT, G-II's ability to hydrolyze G1M9 N-glycan on both native and non-native forms of glycoproteins is essential for preventing prolonged association of glycoproteins in native conformation to lectin chaperone CNX/CRT.

In conclusion, we have synthesized glycoproteins bearing homogeneous G1M9 oligosaccharides in both native and misfolded forms using combined chemical synthesis, enzymatic glucosylation, and iterative intentional misfolding. We found that UGGT can transfer glucose to misfolded dimeric M9-IL-8 in about 90% yield so that we could obtain homogeneous G1M9-IL-8 derivatives. We also found that the concentration of Gn·HCl affects the ratio of misfolded species formation, and we successfully obtained the misfolded G1M9-IL-8 monomer and dimer in sufficient amounts. In vitro assay of G-II revealed that it trims the glucose residue from the G1M9 N-glycan regardless of the protein conformation. Our assay results also suggested that CRT binds to misfolded glycoproteins with the lectin site binding to G1M9 N-glycan and the polypeptide binding site binding to hydrophobic protein surface. This data supports the dual-binding model of CRT. Using these chemically defined glycoprotein probes, we are able to suggest the role of each component in the CNX/CRT cycle: UGGT acts as a primary folding sensor, and CRT interacts with hydrophobic glycoprotein using the dual-binding model, and G-II hydrolyzes everything to move the cycle forward. Our results clearly demonstrate that combined chemical and enzymatic methods give us access to homogeneous Glc_1Man_9-glycoprotein probes with a defined structure, and these probes are versatile chemical tools for gaining insight into the glycoprotein QC system at the molecular level.

Keywords: enzymes · glycopeptides · glycoproteins · mass spectrometry · protein folding

How to cite: *Angew. Chem. Int. Ed.* **2016**, *55*, 3968–3971
Angew. Chem. **2016**, *128*, 4036–4039

[1] a) S. High, F. J. L. Lecomte, S. J. Russell, B. M. Abell, J. D. Oliver, *FEBS Lett.* **2000**, *476*, 38–41; b) D. B. Williams, *J. Cell Sci.* **2006**, *119*, 615–623.

[2] a) C. D'Alessio, J. J. Caramelo, A. J. Parodi, *Semin. Cell Dev. Biol.* **2010**, *21*, 491–499; b) B. R. Pearse, D. N. Hebert, *Biochim. Biophys. Acta Mol. Cell Res.* **2010**, *1803*, 684–693.

[3] a) A. J. Parodi, *Annu. Rev. Biochem.* **2000**, *69*, 69–93; b) A. Helenius, M. Aebi, *Science* **2001**, *291*, 2364–2369; c) A. Helenius, M. Aebi, *Annu. Rev. Biochem.* **2004**, *73*, 1019–1049; d) J. J. Caramelo, A. J. Parodi, *J. Biol. Chem.* **2008**, *283*, 10221–10225.

[4] a) D. N. Hebert, M. Molinari, *Physiol. Rev.* **2007**, *87*, 1377–1408; b) A. Buchberger, B. Bukau, T. Sommer, *Mol. Cell* **2010**, *40*, 238–252.

[5] M. Aridor, *Adv. Drug Delivery Rev.* **2007**, *59*, 759–781.

[6] a) 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; b) J. J. Caramelo, O. A. Castro, G. de Prat-Gay, A. J. Parodi, *J. Biol. Chem.* **2004**, *279*, 46280–46285; c) S. C. Taylor, A. D. Ferguson, J. J. M. Bergeron, D. Y. Thomas, *Nat. Struct. Mol. Biol.* **2004**, *11*, 128–134; d) C. Ritter, K. Quirin, M. Kowarik, A. Helenius, *EMBO J.* **2005**, *24*, 1730–1738.

[7] a) A. Zapun, S. M. Petrescu, P. M. Rudd, R. A. Dwek, D. Y. Thomas, J. J. M. Bergeron, *Cell* **1997**, *88*, 29–38; b) Y. Ihara, M. F. Cohen-Doyle, Y. Saito, D. B. Williams, *Mol. Cell* **1999**, *4*, 331–341.

[8] E. S. Trombetta, A. J. Parodi, *Methods* **2005**, *35*, 328–337.

[9] a) K. Totani, Y. Ihara, I. Matsuo, Y. Ito, *J. Biol. Chem.* **2006**, *281*, 31502–31508; b) K. Totani, Y. Ihara, T. Tsujimoto, I. Matsuo, Y. Ito, *Biochemistry* **2009**, *48*, 2933–2940.

[10] M. N. Amin, W. Huang, R. M. Mizanur, L. X. Wang, *J. Am. Chem. Soc.* **2011**, *133*, 14404–14417.

[11] a) M. Izumi, Y. Makimura, S. Dedola, A. Seko, A. Kanamori, M. Sakono, Y. Ito, Y. Kajihara, *J. Am. Chem. Soc.* **2012**, *134*, 7238–7241; b) S. Dedola, M. Izumi, Y. Makimura, A. Seko, A. Kanamori, M. Sakono, Y. Ito, Y. Kajihara, *Angew. Chem. Int. Ed.* **2014**, *53*, 2883–2887; *Angew. Chem.* **2014**, *126*, 2927–2931; c) Y. Takeda, A. Seko, M. Hachisu, S. Daikoku, M. Izumi, A. Koizumi, K. Fujikawa, Y. Kajihara, Y. Ito, *Glycobiology* **2014**, *24*, 344–350; for reviews see, d) Y. Ito, Y. Takeda, A. Seko, M. Izumi, Y. Kajihara, *Semin. Cell Dev. Biol.* **2015**, *41*, 90–98; e) M. Izumi, S. Dedola, Y. Ito, Y. Kajihara, *Isr. J. Chem.* **2015**, *55*, 306–314.

[12] a) C. Piontek, D. V. Silva, C. Heinlein, C. Pohner, S. Mezzato, P. Ring, A. Martin, F. X. Schmid, C. Unverzagt, *Angew. Chem. Int. Ed.* **2009**, *48*, 1941–1945; *Angew. Chem.* **2009**, *121*, 1974–1978; b) M. Murakami, R. Okamoto, M. Izumi, Y. Kajihara, *Angew. Chem. Int. Ed.* **2012**, *51*, 3567–3572; *Angew. Chem.* **2012**, *124*, 3627–3632; c) A. Reif, S. Siebenhaar, A. Troster, M. Schmalzlein, C. Lechner, P. Velisetty, K. Gottwald, C. Pohner, I. Boos, V. Schubert, S. Rose-John, C. Unverzagt, *Angew. Chem. Int. Ed.* **2014**, *53*, 12125–12131; *Angew. Chem.* **2014**, *126*, 12321–12327; for reviews see, d) C. Unverzagt, Y. Kajihara, *Chem. Soc. Rev.* **2013**, *42*, 4408–4420; e) R. Okamoto, M. Izumi, Y. Kajihara, *Curr. Opin. Chem. Biol.* **2014**, *18*, 92–99.

[13] a) P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. H. Kent, *Science* **1994**, *266*, 776–779; b) E. C. B. Johnson, S. B. H. Kent, *J. Am. Chem. Soc.* **2006**, *128*, 6640–6646.

[14] a) S. P. Thomson, D. B. Williams, *Cell Stress Chaperones* **2005**, *10*, 242–251; b) G. Kozlov, C. L. Pocanschi, A. Rosenauer, S. Bastos-Aristizabal, A. Gorelik, D. B. Williams, K. Gehring, *J. Biol. Chem.* **2010**, *285*, 38612–38620.

[15] a) Y. Saito, Y. Ihara, M. R. Leach, M. F. Cohen-Doyle, D. B. Williams, *EMBO J.* **1999**, *18*, 6718–6729; b) U. G. Danilczyk, D. B. Williams, *J. Biol. Chem.* **2001**, *276*, 25532–25540; c) M. R. Leach, M. F. Cohen-Doyle, D. Y. Thomas, D. B. Williams, *J. Biol. Chem.* **2002**, *277*, 29686–29697.

[16] C. L. Pocanschi, G. Kozlov, U. Brockmeier, A. Brockmeier, D. B. Williams, K. Gehring, *J. Biol. Chem.* **2011**, *286*, 27266–27277.

Received: December 10, 2015
Published online: February 17, 2016

