JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Communication

pubs.acs.org/JACS

Chemical Synthesis of Intentionally Misfolded Homogeneous Glycoprotein: A Unique Approach for the Study of Glycoprotein Quality Control

Masayuki Izumi,[†,‡] Yutaka Makimura,[†,‡] Simone Dedola,[†,‡] Akira Seko,[‡] Akiko Kanamori,[‡,§] Masafumi Sakono,[‡,§] Yukishige Ito,*[,‡,§] and Yasuhiro Kajihara*[,†,‡]

[†]Department of Chemistry, Graduate School of Science, Osaka University, 1-1 Machikaneyama-cho, Toyonaka, Osaka, 560-0043 Japan

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

[§]RIKEN Advanced Science Institute, 2-1 Hirosawa, Wako, Saitama, 351-0198 Japan

Supporting Information

ABSTRACT: Biosynthesis of glycoproteins in the endoplasmic reticulum employs a quality control system, which discriminates and excludes misfolded malfunctional glycoproteins from a correctly folded one. As chemical tools to study the glycoprotein quality control system, we systematically synthesized misfolded homogeneous glycoproteins bearing a high-mannose type oligosaccharide via oxidative misfolding of a chemically synthesized homogeneous glycopeptide. The endoplasmic reticulum folding sensor enzyme, UDP-glucose:glycoprotein glucosyltransferase (UGGT), recognizes a specific folding intermediate, which exhibits a molten globule-like hydrophobic nature.

Glycoproteins, such as cytokines, hormones, or antibodies, play important roles in several biological events.[1] During their biosynthetic pathways, correctly folded glycoproteins are transported to cell surfaces, secreted, or retained in intracellular compartments.[2−4] By contrast, terminally misfolded glycoproteins are delivered to cytosol and eliminated by a degradation process called ER-associated degradation (ERAD). This mechanism is essential in maintaining the vitality of living systems, by maximizing the production of active proteins as well as by avoiding the accumulation of a misfolded glycoprotein, which is inactive and often toxic.

The endoplasmic reticulum (ER) plays central roles in both synthesizing glycoproteins and regulating their folding process.[5] In the ER, numerous chaperones, enzymes, and cargo receptors constitute a "quality control" system, which rigorously discriminates misfolded and correctly folded glycoproteins.[2,3] Among them, lectin chaperones calnexin/calreticulin (CNX/CRT) specifically bind to glycoproteins carrying monoglucosylated high-mannose type glycans, most typically $Glc_1Man_9GlcNAc_2$ (G1M9), and assist in their folding with the aid of noncovalently associated oxidoreductase ERp57. Subsequently, glucosidase II removes the residual glucose residue to convert them to nonglucosylated glycoforms such as $Man_9GlcNAc_2$ (M9). At this stage, incompletely folded glycoproteins are trapped by UDP-glucose:glycoprotein glucosyltransferase (UGGT). UGGT plays a role as the "folding sensor" in the ER by virtue of its ability to recognize incompletely folded glycoproteins.[6] It transfers a glucose unit from UDP-glucose (UDP-Glc) to high-mannose-oligosaccharides (e.g., M9) to form monoglucosylated glycoforms (e.g., G1M9), only when protein folding is immature so that misfolded glycoproteins can engage in iterative interaction with CNX/CRT to maximize their proper folding. Synthetic and nonproteinic substrates of UGGT have previously been discovered,[7,8] which gave a rigorous estimate of the glycan specificity of UGGT. However, the difference in the glycoprotein's reactivity associated with the change in their folding states is yet to be clarified.

In general, protein folding in the ER proceeds sequentially and generates folding intermediates of various conformational states. Whereas previous studies that employed completely denatured glycoproteins[9,10] proposed sophisticated misfolded glycoprotein models,[11−17] they are heterogeneous in glycan structures, and, probably, in their folding states as well. Accordingly, well-characterized homogeneous glycoproteins, in both folded and misfolded forms, are expected to be highly powerful on the basis of UGGT's unique specificity. Due to its dramatic advances in recent years, chemical synthesis is now a realistic option to obtain homogeneous glycoproteins,[18,19] which though may also be produced by using engineered microorganisms[20] or by remodeling of *N*-glycan of natural glycoproteins by a chemoenzymatic approach.[21] In this paper, we describe the very first endeavor to chemically synthesize "stably misfolded" and homogeneous glycoproteins. Their usefulness as chemical tools for exploring the mechanism of substrate specificity of UGGT is demonstrated.

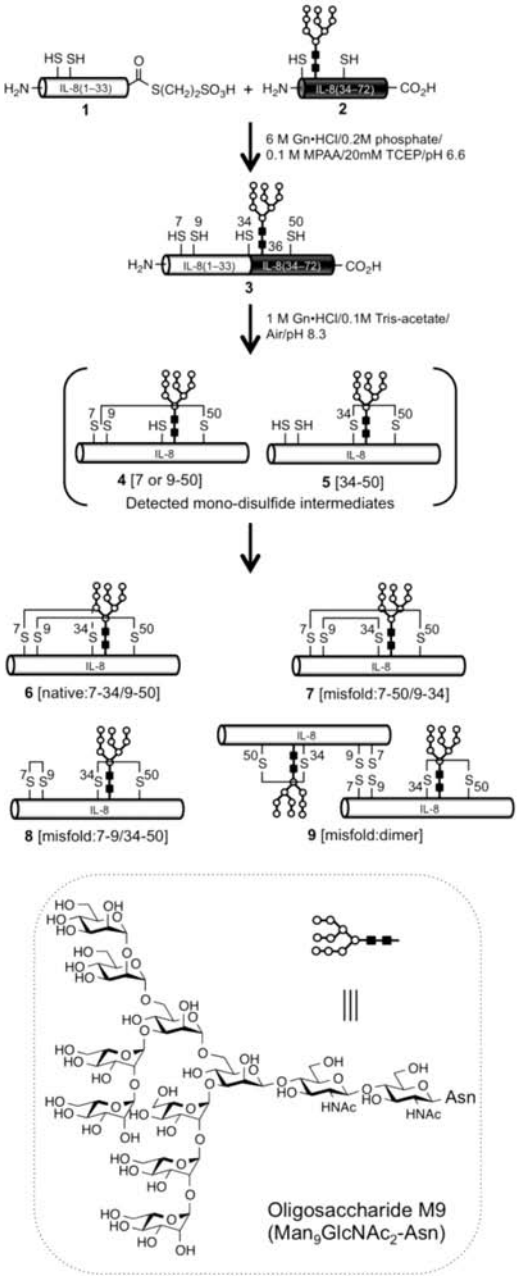
As a model glycoprotein, we chose interleukin-8 (IL-8, CXCL8), although it is not naturally glycosylated. Our decision to incorporate an oligosaccharide to IL-8 was made based on the following: (i) its tertiary structure has been well studied by X-ray crystallography[22] and NMR analysis;[23] (ii) native IL-8 consists of 72 amino acid residues with two disulfides between Cys7-Cys34 and Cys9-Cys50, allowing us to prepare intentionally misfolded proteins by shuffling these linkages; (iii) since chemical synthesis of IL-8 by using native chemical

Received: February 9, 2012
Published: April 12, 2012

   dx.doi.org/10.1021/ja3013177 | *J. Am. Chem. Soc.* 2012, 134, 7238−7241

ligation (NCL) was achieved,[24] all technologies to prepare its polypeptide chain are in our hands. In order to synthesize the glycosylated IL-8, the full-length glycopeptide was divided into two segments **1** and **2**, which correspond to residues 1−33 and 34−72, respectively (Scheme 1, Figure 2a). We planned to

Scheme 1. Synthesis of Native and Misfolded Glycosylated IL-8s Bearing Oligosaccharide M9



incorporate the M9 oligosaccharide to the Asn36 (Figure 2a). Because this position is located on a loop between two β-strands (Figure 2f), we anticipated that it could be modified without perturbing the tertiary structure of IL-8. Preparation of two segments **1** and **2** was achieved by employing a hybrid Boc/Fmoc strategy. Namely, peptide-$^{\alpha}$thioester segment **1** consisting of residue 1−33 was prepared by using an *in situ* neutralization *tert*-Boc SPPS protocol,[25] while glycopeptide segment **2** having an N-terminal cysteine residue was prepared on an HMPA-PEGA resin with Fmoc chemistry.[18] Details of the synthesis are provided in the Supporting Information (SI) text (Figures S1, S2). Subsequent NCL of peptide-$^{\alpha}$thioester **1** and glycopeptide **2** employing standard NCL buffer containing 6 M guanidine hydrochloride, 0.2 M sodium phosphate, 0.1 M 4-mercaptophenylacetic acid (MPAA), and 20 mM tris(2-carboxyethyl)phosphine (TCEP)[26] gave the desired full-length glycopeptide **3** bearing the M9 oligosaccharide in good overall yield (Figure S3).

In order to obtain glycosylated IL-8 in a correctly folded form, oxidative folding of the full-length glycopeptide **3** was examined under redox conditions (1 mM cysteine and 0.05 mM cystine). Progress of the folding reaction was carefully monitored by RP-HPLC, the profile of which is shown in Figure 1a. A few peaks of the folding intermediates were

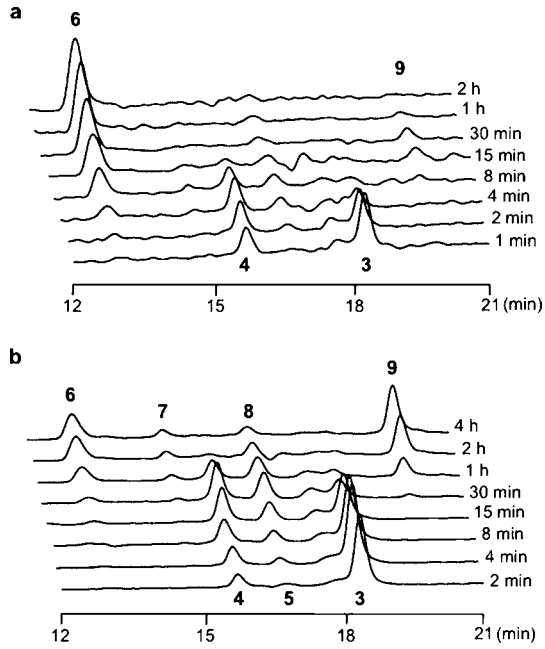


Figure 1. Time-resolved HPLC profiles of oxidative folding of full-length glycopolypeptide **3** in 1 M Gn•HCl, 0.1 M Tris-HCl, pH 8.3. HPLC traces were recorded at 220 nm. (a) Oxidative folding under 1 mM cysteine, 0.05 mM cystine. (b) Oxidative misfolding under air oxidation conditions. The folding buffer was bubbled with air and used immediately. Peak labels indicate the compound numbers in Scheme 1. Compounds **5** and **8** were distinguished by MS.

evident at the early stage, a majority of which converged into a single product after 2 h. The major product **6** was confirmed to have two native disulfide bonds between Cys7-Cys34 and Cys9-Cys50 as expected (see SI text and Figures S5−S15 for disulfide mapping).

A subsequent attempt was made to shunt disulfide reshuffling and to isolate folding intermediates deliberately, through oxidative folding of **3** in the absence of cysteine and cystine. As shown in Figure 1b, after 30 min of reaction, two major products corresponding to the peaks with retention times at 15.7 and 16.7 min were isolated, structures of which were determined to be **4** (disulfide bonds between Cys50-Cys7 and/or Cys50-Cys9) and **5** (Cys50-Cys34), respectively (Scheme 1, Figures S16−S18). After 4 h, four peaks with retention times at 13.0, 14.9, 16.7, and 19.8 min were isolated, which were assigned to be **6** (native), **7** (disulfide bonds between Cys7-Cys50, Cys9-Cys34), **8** (Cys7-Cys9, Cys34-Cys50), and **9**,

dx.doi.org/10.1021/ja3013177 | J. Am. Chem. Soc. 2012, 134, 7238−7241

respectively (Scheme 1, Figures S16, S19, S20). Interestingly, **9** was revealed to be a homodimer, which was assumed to have two interpeptide disulfide linkages between Cys7 and Cys9 of two IL-8 polypeptides, in addition to the intrastrand linkage between Cys34 and Cys50 (Scheme 1, Figure S21).

Since misfolded homogeneous glycoproteins **7**, **8**, and **9** were formed during disulfide formation, they may well mimic non-native conformers formed in the ER. Accordingly, we set out to examine them as substrates of UGGT (Figures 2b, S22, S23). Since one of the bis-disulfides **8** was unstable under the assay conditions, the assay was conducted by using **7** and **9** in comparison with the native-type isomer **6**. Recombinant human UGGT produced in an *E. coli* expression system (SI text) was incubated with **6**, **7**, or **9**, in the presence of UDP-Glc (0.5 mM), and progress of the reaction was monitored by LC-MS. As shown in Figures 2b, S22, and S23, glucose transfer to the native **6** was not observed within MS detection limits, indicating that the amount of formed G1M9-**6** was very small, if any. An increased amount of UGGT or extended incubation gave the same result. On the other hand, the misfolded monomer **7** exhibited activity toward the enzyme, giving ca. 20% of the product after 20 h. Interestingly, the misfolded dimer **9** was strikingly reactive, and the extent of glucosylation reached more than 50% in 2 h. Because the dimer **9** contains two M9 oligosaccharides, the mass spectrum revealed the formation of two UGGT products having one and two glucose residues, respectively (Figures 2c, S23). Even though **6**, **7**, and **9** shared exactly the same peptide and oligosaccharide sequences, UGGT exhibited a clear substrate preference in the order **9** ≫ **7** > **6** (Figure 2b).

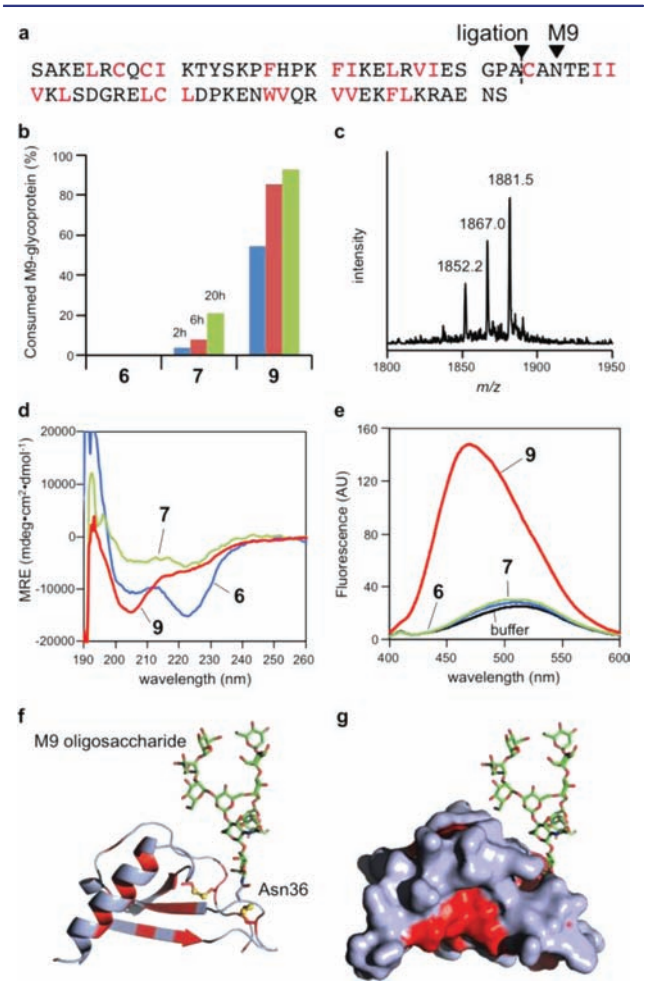


Figure 2. (a) Amino acid sequence of M9-IL-8. Hydrophobic residues were indicated in red using the Kyte–Dolittle hydrophobic scale. Ligation site and glycosylation site were indicated as "ligation" and "M9", respectively. (b) Percentage of consumed M9-IL-8 **6**, **7**, **9** by UGGT reaction after 2 h (blue), 6 h (red), and 20 h (green). (c) Expansion of 11+ region of representative mass spectrum of **9** after reaction with UGGT. 1852.2 (**9**), 1867.0 (**9**+1Glc), 1881.5 (**9**+2Glc). (d) Far-UV CD spectra of **6** (blue), **7** (green), and **9** (red). (e) ANS binding fluorescence spectra of **6** (blue), **7** (green), **9** (red), and buffer (black). (f) Ribbon diagram model of **6** based on PDB(3IL8) with hydrophobic residues in red. M9 oligosaccharide was added to the protein model using glyprot (http://www.glycosciences.de/modeling/glyprot/). (g) Surface model of **6** with hydrophobic residues in red. Both model graphics were prepared using Pymol (http://www.pymol.org/).

To elucidate the protein structure of the preferred substrate of UGGT, we next analyzed the structural features of **6**, **7**, and **9** using far-UV circular dichroism (CD) spectroscopy (Figure 2d). The native isomer **6** had a very similar CD spectrum profile to that reported for the native nonglycosylated IL-8[27] which consists of α-helix and β-sheet structures (Figure 2f), confirming our assumption that the oligosaccharide attached to Asn36 will not affect the three-dimensional structure of the protein. On the other hand, misfolded monomer **7** and the dimer **9** showed markedly different CD spectra, suggesting that they are forced to have a non-native secondary structure. These results indicate that **6**, **7**, and **9** have conformational character different from each other.

Subsequently, their affinities to 1-anilino-8-naphthalenesulfonate (ANS) were evaluated (Figure 2e). ANS is known to be a gauge of the protein surface hydrophobicity, which gives intense fluorescence emission upon binding to a protein. Previous studies have shown that proteins with partially folded molten globule (MG)-like states have the highest affinity to ANS, indicating that hydrophobic patches are maximally exposed in the MG state.[11] As shown in Figure 2e, the dimer **9**, but not **6** (folded monomer) and **7** (misfolded monomer), showed intense emission caused by ANS binding, indicating that **9** had surface exposed hydrophobicity to the largest extent. This assumption was substantiated by their simple C4 RP-HPLC retention times (Figure 1b), suggesting that the overall hydrophobicity order appears to be **9** > **7** > **6**.

These results clearly showed that the extent of reactivity as a substrate of UGGT was in parallel with their affinity to ANS, the order of magnitude being highest for **9** in both measurements. Close correlation between these properties strongly corroborates the proposed propensities of UGGT and ANS, both of which have the highest affinity to MG type conformers.[11] In fact, in its folded state, hydrophobic residues of IL-8 are largely buried inside the protein (Figure 2a, g). Upon misfolding, some of these hydrophobic residues are exposed and form hydrophobic patches on the protein's surface, which are recognized by UGGT. Using chemically synthesized homogeneous glycoproteins as substrates for UGGT and LC-MS for the analysis of the reaction product, our study provided concrete evidence which supports that UGGT has the ability to recognize subtle conformational differences among glycoproteins.

It is reported that fully unfolded RNase B is a substrate of UGGT but folding intermediates, which probably contain incorrect disulfide bonds, were not significantly glucosylated.[28] This work and most previous studies on UGGT employed

naturally occurring glycoproteins. Because of their heterogeneity, in terms of glycoform structures and composition of conformers, conclusions have inevitably been speculative. In this work, we achieved the chemical synthesis of glycosylated IL-8 which was grafted with a homogeneous M9 glycan. Most importantly, it allowed us to obtain the glycoprotein in a correctly folded state as well as in misfolded states, in single conformers. Misfolded glycoproteins were formed compulsorily by intentionally suppressing disulfide bond reshuffling during oxidative folding. Our analysis using homogeneous and systematically prepared misfolded glycoproteins showed that UGGT indeed prefers misfolded over correctly folded glycoproteins, clearly indicating that UGGT has substrate recognition sites for both the nonglucosylated glycan and misfolded region of protein. The results, i.e., UGGT's ability to discriminate between correctly folded **6** and misfolded **7**, **9**, also suggested that we could find ideal model glycoproteins as valuable chemical tools for the study of glycoprotein quality control.

Other misfolded analogs, one of which will contain a consensus N-glycosylation sequence Asn-X-Thr/Ser, or analogs bearing additional high-mannose type oligosaccharides, which will be models of glycoproteins having several N-linked oligosaccharides, will be designed based on these small glycosyl-IL-8 derivatives described here, since misfolded homogeneous natural glycoproteins consisting of over 150 amino acid residues are difficult to design and synthesize.

We anticipate that X-ray analysis of a UGGT-glycoprotein binary complex will give complete picture of the molecular mechanism of unique specificity exhibited by UGGT. To that end, our approach, which has enabled rigorous analysis of UGGT's specificity, will be exceptionally powerful in building a solid foundation. By virtue of its ability to incorporate various glycans and non-natural substituents at will, potential of our approach in drawing clear picture of various events in glycoprotein folding system is obvious. We believe that our future study along this line will enhance our understanding of functions and precise recognition mechanism of enzymes, lectins and chaperones involved in the quality control of glycoproteins, such as CNX/CRT, glucosidase II, ERp57, the ER cargo receptors and degradation enzymes, in addition to UGGT.[29,30]

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

General methods, spectroscopic characterization, preparation of recombinant UGGT and $Man_9GlcNAc_2$-(Fmoc-Asn), synthesis of all glycoproteins, disulfide bond mapping of **4**–**9**, and UGGT assay. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

kajihara@chem.sci.osaka-u.ac.jp; yukito@riken.jp

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Dr. Osamu Kanie and Dr. Yoichi Takeda for their valuable advice and Ms. Satoko Shirahata for technical assistance.

■ REFERENCES

(1) Dwek, R. A. *Chem. Rev.* **1996**, *96*, 683.
(2) Helenius, A.; Aebi, M. *Annu. Rev. Biochem.* **2004**, *73*, 1019.
(3) Parodi, A. J. *Annu. Rev. Biochem.* **2000**, *69*, 69.
(4) Takeda, Y.; Totani, K.; Matsuo, I.; Ito, Y. *Curr. Opin. Chem. Biol.* **2009**, *13*, 582.
(5) Malhotra, J. D.; Kaufman, R. J. *Semin. Cell Dev. Biol.* **2007**, *18*, 716.
(6) Trombetta, E. S.; Parodi, A. J. *Methods* **2005**, *35*, 328.
(7) Totani, K.; Ihara, Y.; Matsuo, I.; Koshino, H.; Ito, Y. *Angew. Chem., Int. Ed.* **2005**, *44*, 7950.
(8) Totani, K.; Ihara, Y.; Tsujimoto, T.; Matsuo, I.; Ito, Y. *Biochemistry* **2009**, *48*, 2933.
(9) Parker, C. G.; Fessler, L. I.; Nelson, R. E.; Fessler, J. H. *EMBO J.* **1995**, *14*, 1294.
(10) Sousa, M. C.; Ferrero-Garcia, M. A.; Parodi, A. J. *Biochemistry* **1992**, *31*, 97.
(11) Caramelo, J. J.; Castro, O. A.; Alonso, L. G.; de Prat-Gay, G.; Parodi, A. J. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 86.
(12) Caramelo, J. J.; Castro, O. A.; de Prat-Gay, G.; Parodi, A. J. *J. Biol. Chem.* **2004**, *279*, 46280.
(13) Ritter, C.; Helenius, A. *Nat. Struct. Biol.* **2000**, *7*, 278.
(14) Ritter, C.; Quirin, K.; Kowarik, M.; Helenius, A. *EMBO J.* **2005**, *24*, 1730.
(15) Sousa, M.; Parodi, A. J. *EMBO J.* **1995**, *14*, 4196.
(16) Taylor, S. C.; Ferguson, A. D.; Bergeron, J. J. M.; Thomas, D. Y. *Nat. Struct. Mol. Biol.* **2004**, *11*, 128.
(17) Trombetta, E. S.; Helenius, A. *J. Cell Biol.* **2000**, *148*, 1123.
(18) Yamamoto, N.; Takayanagi, Y.; Yoshino, A.; Sakakibara, T.; Kajihara, Y. *Chem.—Eur. J.* **2007**, *13*, 613.
(19) (a) Yamamoto, N.; Tanabe, Y.; Okamoto, R.; Dawson, P. E.; Kajihara, Y. *J. Am. Chem. Soc.* **2008**, *130*, 501. (b) Murakami, M.; Okamoto, R.; Izumi, M.; Kajihara, Y. *Angew. Chem., Int. Ed.* **2012**, *51*, 3567. (c) Sakamoto, I.; Tezuka, K.; Fukae, K.; Ishii, K.; Taduru, M.; Maeda, M.; Ouchi, K.; Yoshida, Y.; Nambu, Y.; Igarashi, J.; Hayashi, N.; Tsuji, T.; Kajihara, Y. *J. Am. Chem. Soc.* **2012**, *134*, 5428.
(20) Rich, J. R.; Withers, S. G. *Nat. Chem. Biol.* **2009**, *5*, 206 and references cited therein.
(21) Amin, M. N.; Huang, W.; Mizanur, R. M.; Wang, L. X. *J. Am. Chem. Soc.* **2011**, *133*, 14404.
(22) Baldwin, E. T.; Weber, I. T.; Charles, R., St; Xuan, J. C.; Appella, E.; Yamada, M.; Matsushima, K.; Edwards, B. F.; Clore, G. M.; Gronenborn, A. M. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 502.
(23) Clore, G. M.; Appella, E.; Yamada, M.; Matsushima, K.; Gronenborn, A. M. *Biochemistry* **1990**, *29*, 1689.
(24) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. H. *Science* **1994**, *266*, 776.
(25) Schnölzer, M.; Alewood, P.; Jones, A.; Alewood, D.; Kent, S. B. H. *Int. J. Pept. Res. Ther.* **2007**, *13*, 31.
(26) Johnson, E. C. B.; Kent, S. B. H. *J. Am. Chem. Soc.* **2006**, *128*, 6640.
(27) Goger, B.; Halden, Y.; Rek, A.; Mösl, R.; Pye, D.; Gallagher, J.; Kungl, A. J. *Biochemistry* **2002**, *41*, 1640.
(28) Zapun, A.; Petrescu, S. M.; Rudd, P. M.; Dwek, R. A.; Thomas, D. Y.; Bergeron, J. J. M. *Cell* **1997**, *88*, 29.
(29) Aebi, M.; Bernasconi, R.; Clerc, S.; Molinari, M. *Trends Biochem. Sci.* **2010**, *35*, 74.
(30) Lederkremer, G. Z. *Curr. Opin. Struct. Biol.* **2009**, *19*, 515.