Supporting Information of

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

*Corresponding Author. E-mail:kajihara@chem.sci.osaka-u.ac.jp; yukito@riken.jp

Supplementary Methods

General and abbreviation. 2-(*1H*-Benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU), 1-hydroxy-benzotriazole (HOBt), Boc-amino acids and Fmoc-amino acids were purchased from Peptide Institute Inc. Boc-Cys(Trt) was purchased from Watanabe Chemical Ind. Boc-Leu-OCH_2-Pam resin was purchased from AnaSpec Inc. S-Trityl-mercaptopropionic acid was purchased from Oakwood Products Inc. Trifluoroacetic acid (TFA), sodium 2-mercaptoethanesulfonate (MESNa), N-methylimidazole, N,N'-diisopropylcarbodiimide (DIC), N,N-diisopropylethylamine (DIEA), 1,2-ethanedithiol (EDT), triisopropylsilane (TIPS), tris(2-carboxyethyl)phosphine hydrochloride (TCEP) were purchased from Tokyo Chemical Ind. N,N-dimethylformamide (DMF), dichloromethane (DCM), m-cresol, thioanisole, trifluoromethanesulfonic acid (TFMSA), diethyl ether, endopeptidase Lys-C, D,L-dithiothreitol (DTT), N-ethylmaleimide (NEM) were purchased from Wako Pure Chemical. Boc-Arg(di-Z), 1-(mesitylene-2-sulfonyl)-3-nitro-*1H*-1,2,4-trizole (MSNT), amino-PEGA resin, 4-hydroxymethylphenylacetic acid (HMPA) were purchased from Novabiochem. N-Methyl-2-pyrrolidinone (NMP), Guanidine hydrochloride (Gn•HCl), and HPLC grade acetonitrile were purchased from Kanto Chemical. 3-(Diethoxyphosphoryloxy)-1,2,3-benzotriazin-4(*3H*)-one (DEPBT), 4-mercaptophenylacetic acid (MPAA) were purchased from Sigma-Aldrich. RP-HPLC analyses were performed on an Agilent 1260 HPLC system equipped with a multi wavelength detector. Eclipse plus C18 (Agilent Technologies) or Vydac C18, C4 (GRACE Co.) were used for analytical or/and semipreparative HPLC. Cadenza CD-C18 (Imtakt) or Proteonavi (Shiseido) were used in LC/MS analysis. LC/MS spectra were recorded on a Bruker Daltonics amaZon-SL mass spectrometer equipped with an Agilent 1260 HPLC system.

Spectroscopic characterization (ANS and CD spectra) of native and misfolded M9-IL-8 glycoproteins. Far-UV CD spectra were measured with a JASCO-J820 CD spectropolarimeter. Concentrations of glycoproteins were 2.8 μM for **6**, 1.5 μM for **7**, and 1.4 μM for **9**. Binding of 8-anilinonaphthalene-1-sulfonic acid (ANS)[1] was measured with a JASCO FP-6500 fluorescence spectrometer. Each glycoprotein (1

μM) was incubated with ANS (50 μM) at 37ºC for 10 min. The excitation wavelength was set at 360 nm and the emission was scanned from 400 to 600 nm in the correct spectrum mode. All measurements were carried out in 50 mM Tris-HCl, 5 mM $CaCl_2$, pH 7.5.

Preparation of recombinant UDP-glucose:glycoprotein glucosyltransferase (UGGT)

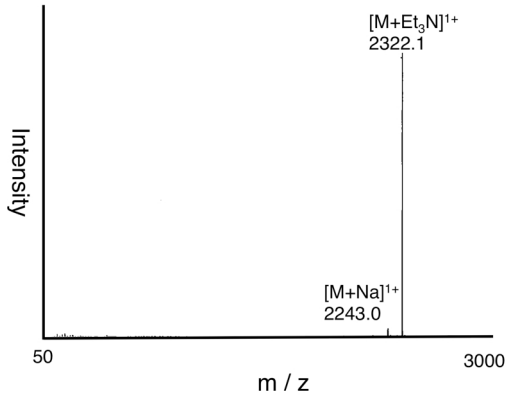
The cDNA fragment encoding truncated form of human UGGT1 (37-1555 aa) that lacked the secretory signal sequence, was cloned into pCold I (Takara Bio Inc., Otsu, Japan) between EcoRI and XbaI restriction sites, and introduced into *Escherichia coli* BL21 cells.

Recombinant UGGT was prepared according to the manufacturer's instructions. Briefly, transformed cells at log phase in 1.5 L of LB broth containing 100 μg/ml ampicillin were cooled at 15 °C for 30 min. Isopropyl β-D-1-thiogalactopyranoside was added at a final concentration of 0.1 mM and induction was performed at 15 °C for 16 h. The cells were collected and sonicated in 200 ml of 20 mM potassium phosphate buffer (pH 8.0), 0.3 M NaCl, and 20 mM imidazole (buffer A) containing 1 mM PMSF, 1 μg/ml pepstatin, 1μg/ml leupeptin, and 3 mM 2-mercaptoethanol, followed by addition of 4 ml of 10% Triton X-100 (v/v). After centrifugation, the supernatant was added 0.2 ml of Ni-NTA agarose (QIAGEN GmbH,Hilden, Germany), and the mixture was allowed to stand for 30 min with occasional mixing. The resin was packed into a plastic column and washed with 30 ml of buffer A. Thereafter, the recombinant protein was eluted with buffer A containing 0.25 M imidazole. The eluate was dialyzed against PBS and concentrated with an Amicon Ultra centrifugal filter 10K (Millipore Corp., Billerica, MA). The yield of the protein was approximately 40 μg.

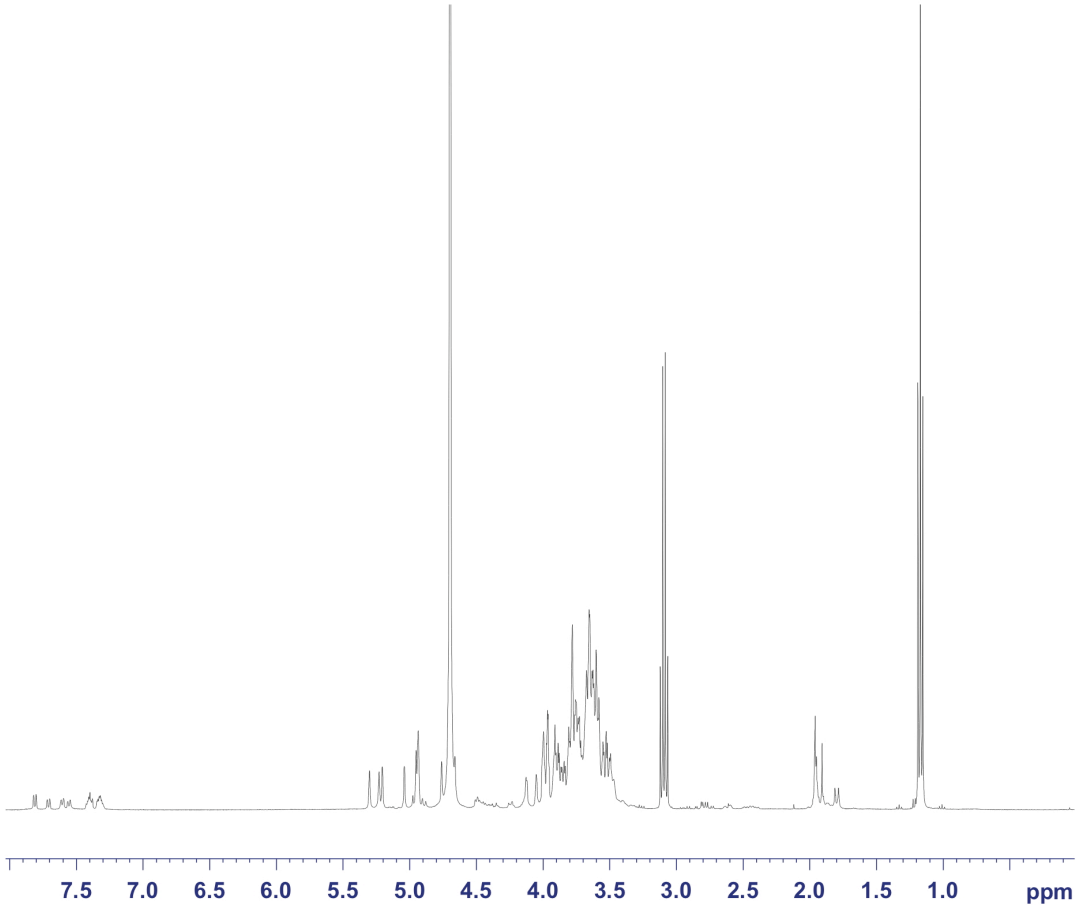
Preparation of $Man_9GlcNAc_2$-(Fmoc-Asn). We have previously reported[2] the preparation of Fmoc-Asn-linked complex-type oligosaccharide from delipidated egg yolk[3] by employing the reported isolation strategy.[4] It has been reported that oligosaccharide Man9GlcNAc2 also exists in egg yolk.[5] Therefore we applied our preparation method of Fmoc-Asn-linked complex-type oligosaccharides to prepare

$Man_9GlcNAc_2$-(Fmoc-Asn). Detailed procedure will be reported elsewhere.

Briefly, delipided hen egg yolk was partialy hydrolyzed by protease from *Bacillus stearothermophilus* (Protin PD-10F Amano enzyme Co. Ltd) and the resulting glycopeptide rich fraction was fully hydrolyzed by protease from *Streptomyces griseus* (Pronase, Roche Co. Ltd.) to oligosaccharyl-asparagine. After gel filtration (sephadex G-50), oligosaccharide-Asn rich fraction was treated with Fmoc-OSu in aqueous DMF containing $NaHCO_3$. Fmoc protected highmannose-type oligosaccharide-Asn was collected by reverse phase column chromatography (Cosmosil 75C18-OPN). Purification by normal phase HPLC column with isocratic conditions (0.1 M AcOH-Et_3N:CH_3CN = 1:1) gave $Man_9GlcNAc_2$-(Fmoc-Asn) fraction. The desired $Man_9GlcNAc_2$-(Fmoc-Asn) was obtained after gel filtration (Sephadex G25) and ion-exchange column (Dowex 50W-X8), which was directly used in solid phase glycopeptide synthesis. ^1H NMR (D_2O) δ 1.81 (s, 3H, GlcN*Ac*), 1.96 (s, 3H, GlcN*Ac*), 3.47–4.12 (br m, 62H, H-2,3,4,5,6 of Man and GlcNAc), 4.49 (d, 1H, H-1, GlcNAc), 4.66 (s, 1H, H-1, Man), 4.76 (s, 1H, H-1, Man), 4.89 (d, 1H, H-1, GlcNAc), 4.94 (2s, 2H, H-1, Man), 4.95 (s, 1H, H-1, Man), 5.04 (s, 1H, H-1, Man), 5.20 (s, 1H, H-1, Man), 5.23 (s, 1H, H-1, Man), 5.30 (s, 1H, H-1, Man), 7.33 (m, 2H, Fmoc), 7.40 (m, 2H, Fmoc), 7.55 (d, J = 7.2 Hz, 1H, Fmoc), 7.60 (d, J = 7.2 Hz, 1H, Fmoc), 7.70 (d, J = 7.4 Hz, 1H, Fmoc), 7.81 (d, J = 7.3 Hz, 1H, Fmoc); ESI-MS calcd for $C_{89}H_{134}N_4O_{60}$ (2218.76); found for *m/z* 2243.0 $[M+Na]^+$ and 2322.10 $[M+Et_3N]^+$.



ESI-MS spectrum of $Man_9GlcNAc_2$-(Fmoc-Asn)



^{1}H NMR spectrum of $Man_9GlcNAc_2$-(Fmoc-Asn)

Supplementary Results

Solid phase peptide synthesis (SPPS) of N-terminal peptide-$^{\alpha}$thioester (1) and C-terminal glycopeptide (2). Peptide-$^{\alpha}$thioester segment **1** consists of residue 1–33 was prepared using *in situ* neutralization Boc SPPS protocol[6] on TAMPAL-OCH_2-Pam resin.[7] Deprotection and cleavage from the resin was performed using high-acid TFMSA conditions.[8] Then, the crude peptide-$^{\alpha}$thioester was treated with MESNa to remove a 2,4-dinitrophenyl group from histidine residue to yield 33-residue MESNa thioester **1** (Fig. S1). Glycopeptide segment **2** having N-terminal cysteine residue was prepared on HMPA-PEGA resin with Fmoc chemistry.[9] Peptide corresponding to residue 37–72 was prepared by microwave assisted coupling and $Man_9GlcNAc_2$-(Fmoc-Asn) was condensed to the peptide. Addition of two more amino acids under diluted conditions yielded the desired glycopeptide **2** bearing M9 oligosaccharide (Fig. S2).

Synthesis of IL-8(1-33)-α-thioester derivative (1). Trityl-associated mercaptopropionic acid-leucine (TAMPAL) Pam resin was prepared according to the literature[7] from Boc-Leu-OCH_2-Pam resin (0.1 mmol) and S-trityl-mercaptopropionic acid. Trityl group was removed by twice 1 min-treatment of TAMPAL-Pam resin with 2.5% TIPS, 2.5% water in TFA. Boc-Ala (75.7 mg, 0.4 mmol) was preactivated with HBTU (0.5 M in DMF, 0.72 mL, 0.36 mmol) and DIEA (0.9 M in DMF, 1.28 mL, 1.15 mmol) for 2 min then the solution was poured onto the resin and the suspension was gently shaken for 1 h at r.t. After the coupling of the first amino acid to the TAMPAL-Pam resin, the segment 1 peptide was elongated using the standard *in situ* neutralization peptide-coupling protocol.[6] For the Boc deprotection, the resin was treated with neat TFA for 1 min × 2. Peptide coupling was carried out with Boc-AA (0.4 mmol) preactivated with HBTU (0.5 M in DMF, 0.72 mL, 0.36 mmol) and DIEA (0.9 M in DMF, 1.28 mL, 1.15 mmol) for 2 min and coupling for 16 min at r.t. Boc-AA used were Pro, Gly, Ser(Bzl), Glu(OBzl), Ile, Val, Arg(di-Z), Leu, Lys(Cl-Z), Phe, His(DNP), Tyr(Br-Z), Thr(Bzl), Cys(MBzl), Gln, Ala. After the coupling of the last amino acid, the resin was washed with DMF and DCM and dried in vacuo. The dried resin (~21 μmol) was treated

with a solution of TFA (4 mL), m-cresol (0.4 mL), thioanisole (0.4 mL) at 0 ºC and TFMSA (0.4 mL) was added dropwise to the suspension and the resulting suspension was vigorously shaken for 2.5 h at 0 ºC. The suspension was filtered and the filtrate was poured into ice-cold ether to form a precipitate. The precipitate was collected by centrifuge and washed twice with ether and air-dried. Then the precipitate was dissolved in a solution of 6 M Gn•HCl, 0.2 M phosphate 0.2 M MESNa, pH 6.4 (3 mL) for 3 h at r.t. The reaction was monitored by HPLC (Eclipse plus C18 Φ4.6 × 100 mm, 0.1% TFA : 0.1% TFA in 90% MeCN = 80 : 20 to 50 : 50 over 15 min at 1 mL/min). The solution was diluted with 8% aqueous TFA and purified by preparative HPLC (Vydac C18 Φ10 × 250 mm, 0.1% TFA : 0.1% TFA in 90% MeCN = 90 : 10 to 80 : 20 over 10 min then 80 : 20 to 55 : 45 over 120 min at 2 mL/min). Fractions containing the desired product were collected and lyophilized to give thioester **1** (8.2 mg, 2.1 μmol, 10 %) as a white foam. ESI-MS: *m/z* calcd. for $C_{174}H_{284}N_{47}O_{48}S_4$ $[M+H]^+$ 3929.7, found 3929.7 (deconvoluted).

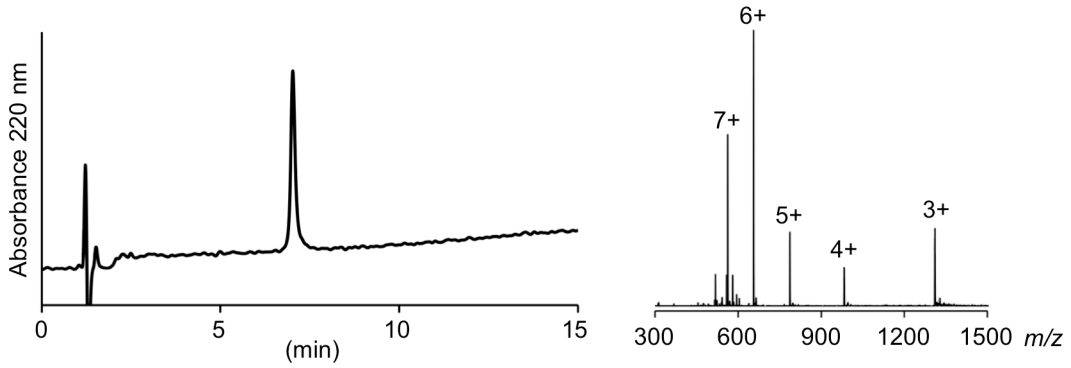


Figure S1. HPLC profile and ESI-MS spectrum of purified **1**.

Synthesis of glycopeptide segment (M9-IL-8: 34-72) (2). HMPA–PEGA resin was prepared from amino-PEGA resin and HMPA as reported previously.[10] Fmoc-Ser(tBu) (95.9 mg, 0.25 mmol) was coupled twice to the resin (~50 μmol) in the presence of MSNT (74.1 mg, 0.25 mmol) and N-methylimidazole (14.9 μL, 0.27 mmol) in DCM (1.25 mL) for 1 h to ensure quantitative coupling. The resin was treated with 20% piperidine in DMF for 15 min at ambient temperature for the removal of Fmoc group. Peptide coupling was carried out using the microwave-assisted protocol as reported in the literature.[11] Briefly, Fmoc-AA (0.32 mmol) was preactivated by treatment with DIC (49.2 μL, 0.32 mmol) and HOBt (43 mg, 0.32 mmol) in NMP (1.59 mL) for 2 min and the

solution was poured onto the resin and the suspension was activated for 20 min by microwave. Fmoc-AA used were Asn(Trt), Glu(OtBu), Ala, Arg(Pbf), Lys(Boc), Leu, Phe, Val, Gln, Trp(Boc), Pro, Asp(OtBu), Gly, Ser(tBu), Ile, Thr(tBu). Coupling of Fmoc-Cys(Trt) was carried out at ambient temperature for 60 min using the same reagents to avoid the possible epimerization. Coupling of Fmoc-Asn(M9) was carried out at about 1 μmol scale using DEPBT and DIEA in DMSO–NMP (1/4, v/v) for 16 h at ambient temperature under protection from light. After the coupling of oligosaccharide-amino acid, Fmoc-Ala (5 equiv) was coupled using DIC, HOBt in NMP twice for 20 min at 45ºC, then Boc-Cys(Trt) (5 equiv) was coupled three times for 60 min at ambient temperature. The resin was treated with 2.5% water, 2.5% EDT, 1% TIPS in TFA for 2 h at ambient temperature. The resin was filtered off and the filtrate was poured into ice-cold ether. A precipitate formed was collected by centrifuge and the filtrate was washed twice with ice-cold ether and air-dried. The precipitate was analyzed by HPLC (Vydac C18 Φ4.6 × 250 mm, 0.1% TFA : 0.1% TFA in 90% MeCN = 70 : 30 to 40 : 60 over 30 min at 1 mL/min) and purified by preparative HPLC (Vydac C18 Φ10 × 250 mm, 0.1% TFA : 0.1% TFA in 90% MeCN = 72 : 28 to 50 : 50 over 110 min at 2 mL/min). Fractions containing the desired product was collected and lyophilized to give glycopeptide **2** as a white foam.; ESI-MS: *m/z* calcd. for $C_{267}H_{442}N_{59}O_{116}S_2$ $[M+H]^+$ 6398.9, found 6398.5 (deconvoluted).

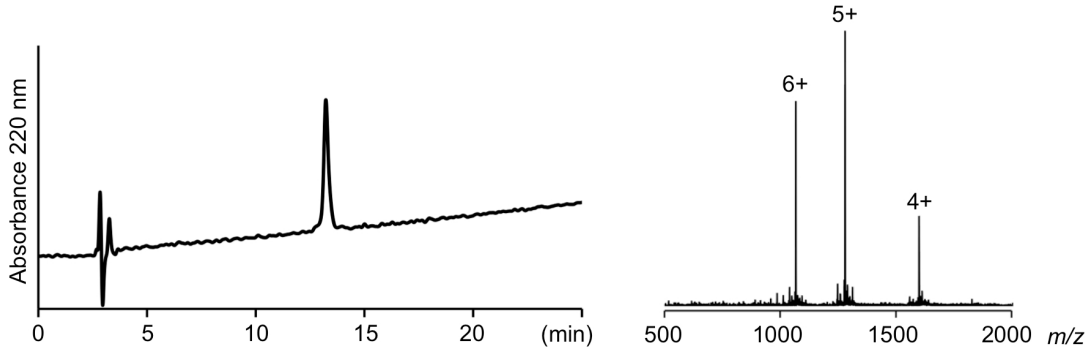


Figure S2. HPLC profile and ESI-MS spectrum of purified **2**.

Synthesis of full-length M9-IL-8 glycopeptide (3). Ligation buffer containing 6 M Gn•HCl, 0.2 M phosphate, 0.1 M MPAA, 20 mM TCEP, pH 6.6 was freshly bubbled with Ar for 2 min. Compound **1** (2.0 mg, 0.5 μmol) and **2** (2.5 mg, 0.4 μmol) were dissolved in the ligation buffer (195 μL) and the solution was stand at room temperature for 15 h. The reaction was monitored by HPLC (Vydac C4 Φ4.6 × 250 mm, 0.1% TFA : 0.1% TFA in 90% MeCN = 70 : 30 to 50 : 50 over 30 min at 1 mL/min). The solution was treated with 10% aqueous DTT (20 μL) for 1 h, then diluted with 4% aqueous TFA (1 mL), and purified by prep HPLC (Vydac C4 Φ10 × 250 mm, 0.04% TFA : 0.04% TFA in 90% MeCN = 75 : 25 to 55 : 45 over 100 min at 2 mL/min). Fractions containing the desired product was collected and lyophilized to give full-length glycopeptide **3** (3.8 mg) as a white foam.; ESI-MS: *m/z* calcd. for $C_{439}H_{719}N_{106}O_{161}S_4$ $[M+H]^+$ 10186.4, found 10185.9 (deconvoluted).

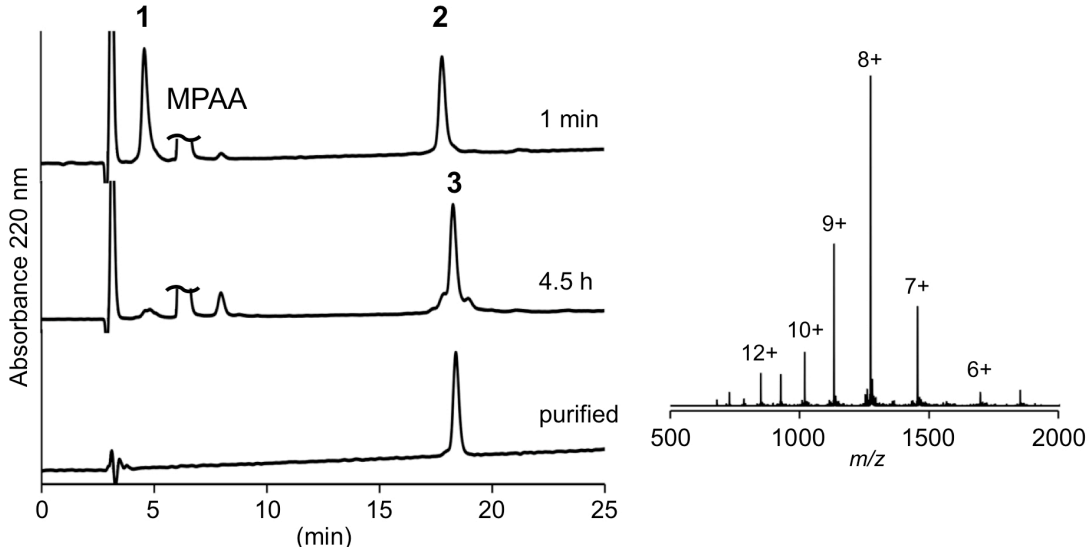


Figure S3. Progress of NCL between **1** and **2** monitored by HPLC and ESI-MS spectrum of purified **3**.

Oxidative folding of M9-IL-8 full-length glycopeptide (3) under cysteine–cystine redox system. Full-length M9-IL-8 glycopeptide **3** was dissolved in a solution of 1 M Gn•HCl, 0.1 M Tris–acetate, 1 mM cysteine, 0.05 mM cystine, pH 8.3 at the concentration of 0.2 mg/mL. In order to monitor folding process, an aliquot at each defined time was diluted with 1/10 amount of 1 M HCl to stop the reaction and then analyzed by HPLC (Vydac C4 Φ4.6 × 250 mm, 0.1% TFA : 0.1% TFA in 90% MeCN = 70 : 30 to 50 : 50 over 30 min at 1 mL/min) and LC/MS (Proteonavi Φ2.0 × 150 mm, 0.1% FA : 0.1% FA in 90% MeCN = 75 : 25 to 55 : 45 over 20 min at 0.2 mL/min). After 4 h, the solution was directly purified by HPLC (Vydac C4 Φ4.6 × 250 mm, 0.04% TFA : 0.04% TFA in 90% MeCN = 75 : 25 for 10 min then 75 : 25 to 55 : 45 over 100 min at 1 mL/min) and lyophilized to give native form of glycosylated IL-8 **6**. This folding process is shown in Figure 1a; ESI-MS: *m/z* calcd. for $C_{439}H_{715}N_{106}O_{161}S_4$ $[M+H]^+$ 10182.4, found 10181.8 (deconvoluted).

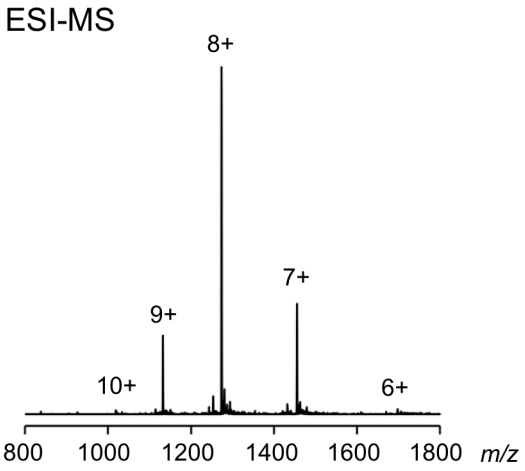
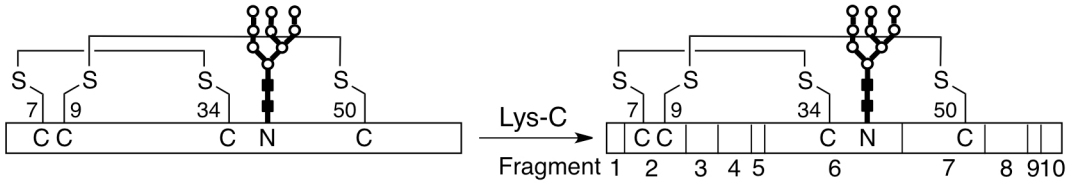


Figure S4. ESI-MS spectrum of purified **6**.

Determination of disulfide linkages in native form of M9-IL-8 (6).

In order to determine disulfide bond positions in **6**, we examined sequential procedures consist of 1) partial reduction of disulfide bond with TCEP, 2) alkyl labeling of free thiol thus reduced with NEM, 3) peptidase digestion with endopeptidase Lys-C and 4) complete reduction of remaining disulfide bond with DTT and MS/MS analysis of the reduced fragments were examined (Fig. S7–S15).

Prior to this experiment, peptidase digestion positions were confirmed by use of endopeptidase Lys-C (Fig. S5, S6). M9-IL-8 **6** in 0.1 M Tris-HCl (pH 6.5) was treated with endopeptidase Lys-C at 37ºC for 30 min. Cysteine-containing fragments (2, 6, 7) were analyzed by LC-MS (cadenza CD-C18 Φ2.0 × 100 mm, 0.1% FA : 0.1% FA in 90% MeCN = 95 : 5 to 40 : 60 over 20 min at 40ºC at 0.2 mL/min).



Fragment 1:SAK
Fragment 2:ELRCQCIK
Fragment 3:TYSK
Fragment 4:PFHPK
Fragment 5:FIK
Fragment 6:ELRVIESGPACANTEIIVK
Fragment 7:LSDGRELCLDPK
Fragment 8:ENWVQRVVEK
Fragment 9:FLK
Fragment 10:RAENS

Figure S5. Amino acid sequences of peptide fragments generated by endopeptidase Lys-C digestion of M9-IL-8. The numbers shown under the peptide cartoon (right) mean fragments number generated by endopeptidase Lys-C. Cysteine residues were indicated in green and a glycosylated asparagine residue was indicated in blue.

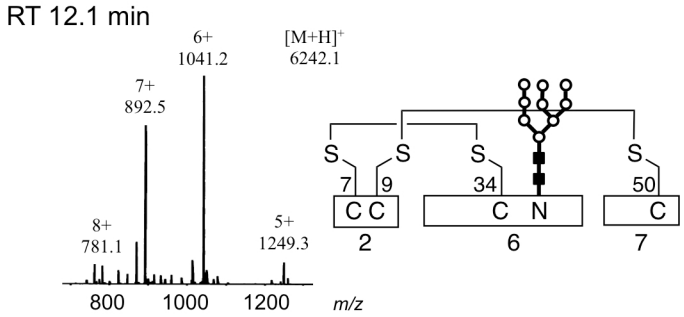


Figure S6. LC/MS spectrum of cysteine-containing fragments (2, 6, 7) after Lys-C digestion of **6**. MS spectrum suggested that three cysteine-containing fragments (2, 6, 7) were all linked together through disulfide bonds.

From this result, pairs of cysteines involved in disulfide bonds were expected to be (7–34, 9–50) or (7–50, 9–34). To determine which disulfide bond pair is formed in **6**, four-step analysis mentioned above was carried out on M9-IL-8 (**6**).

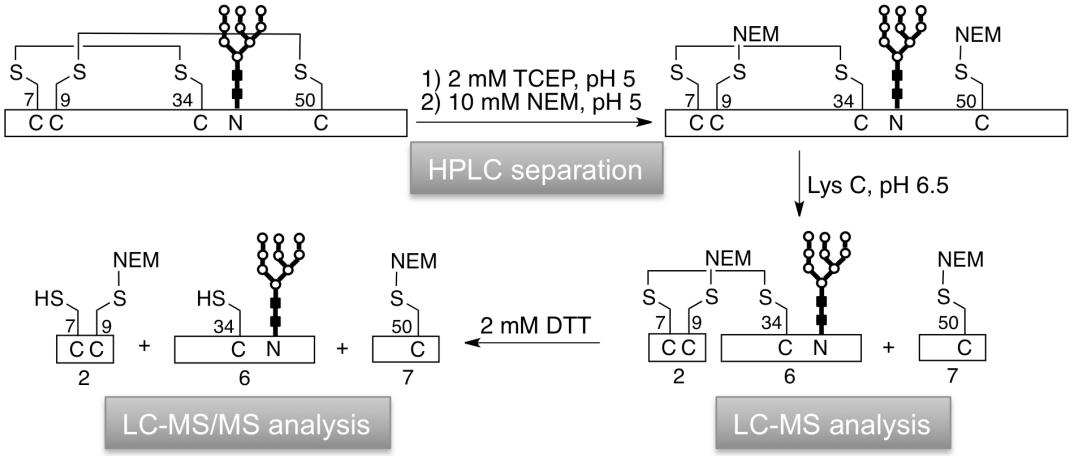


Figure S7. Scheme for the determination of disulfide bond pairs in M9-IL-8 **6**.

Partial reduction and alkyl labeling of 6.

M9-IL-8 (**6**) in 0.1 M Tris-HCl (pH 6.5) was added 20 mM TCEP in water to a final concentration of 2 mM. The pH of the solution was about 5. The solution was allowed to react for 20 min at 37ºC. To the solution was added 50 mM *N*-ethylmaleimide (NEM) in the buffer (pH 6.5) and final concentration of NEM was adjusted to 10 mM. The solution was kept at 37ºC for 30 min. The solution was diluted with the same amount of 4% TFA and the resultant derivatives were separated by HPLC (Vydac C4 Φ4.6 × 250 mm, 0.1% TFA : 0.1% TFA in 90% MeCN = 70 : 30 to 50 : 50 over 1 h at 1 mL/min). The fractions containing derivatives were collected and then lyophilized. The results are shown in Figure S8 and S9.

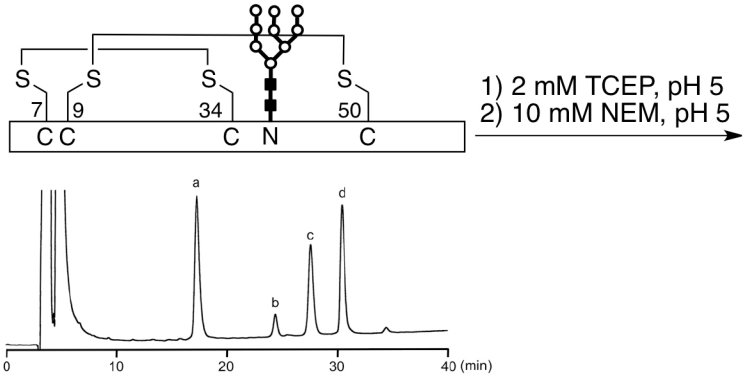


Figure S8. HPLC profile of the two-step reaction (partial reduction and alkyl labeling) of **6**.

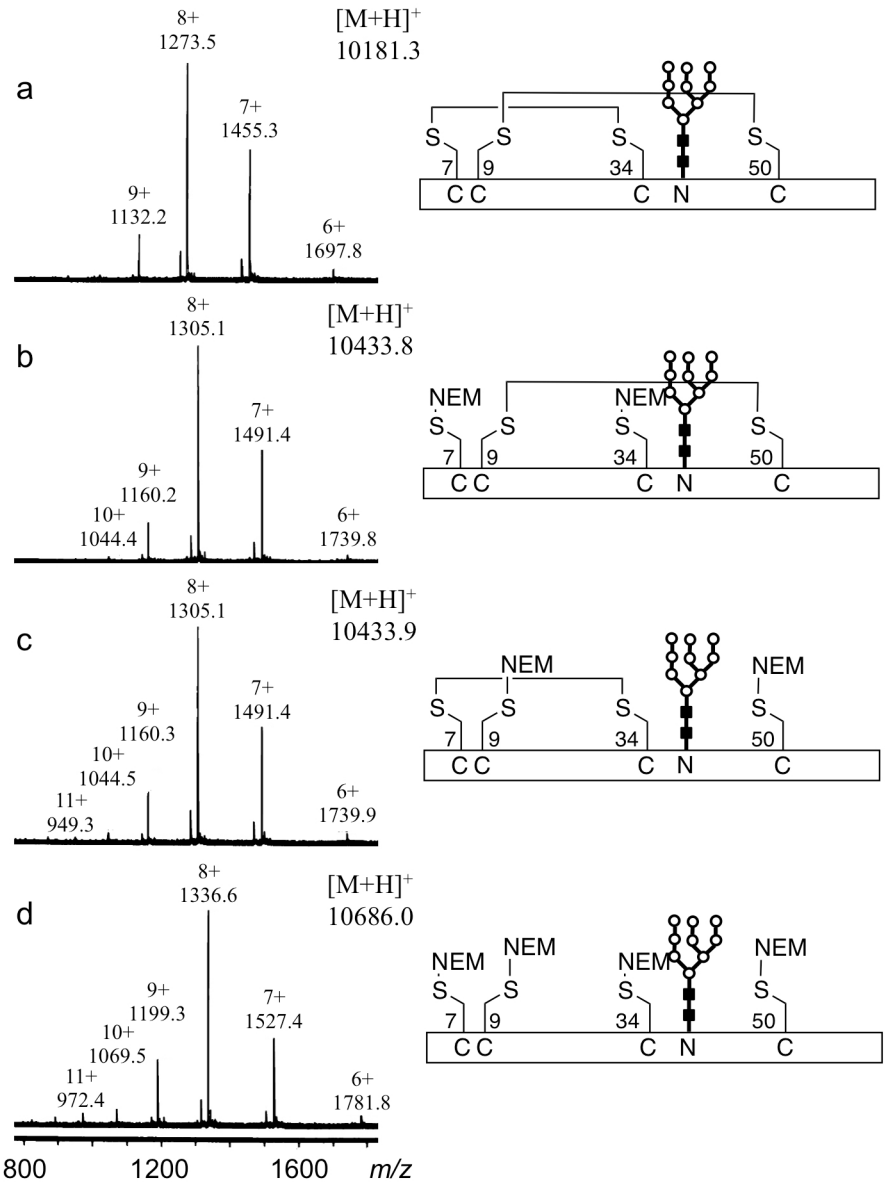


Figure S9. LC/MS spectra of peaks a–d of Figure S8. Peak a) **6**. Peak b) and c) one disulfide bond with two NEMs. Peak d) fully reduced M9-IL-8 with four NEMs. These structures were determined by following endopeptidase Lys-C digestion.

Endopeptidase Lys-C digestion and DTT reduction of the peak c of Figure S8.

Compound corresponds to peak c shown in Figure S8 was treated with Lys-C in 0.1 M Tris-HCl pH 6.5 for 30 min at 37ºC. An aliquot was treated with 4% aqueous TFA and analyzed by LC/MS (cadenza CD-C18 Φ2.0 × 100 mm, 0.1% FA : 0.1% FA in 90% MeCN = 95 : 5 to 40 : 60 over 20 min at 40ºC at 0.2 mL/min). Remaining solution was further treated with DTT at the final concentration of 2 mM for 1h at r.t. and an aliquot was analyzed by LC/MS/MS using the same LC conditions.

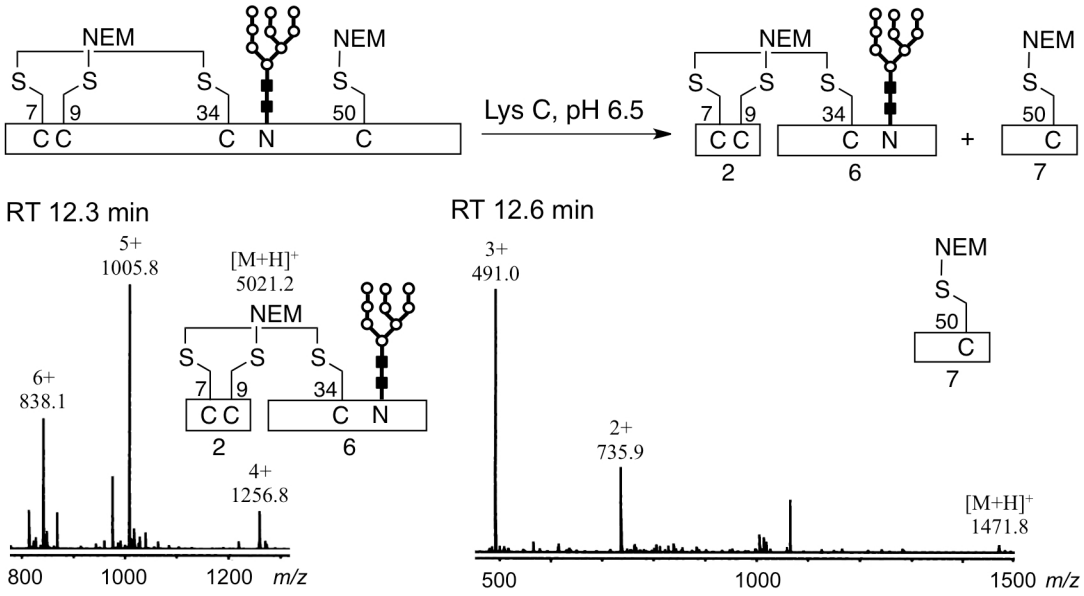


Figure S10. LC/MS spectra of cysteine-containing fragments (2, 6, 7) of the peak c shown in Figure S8 after Lys-C digestion.

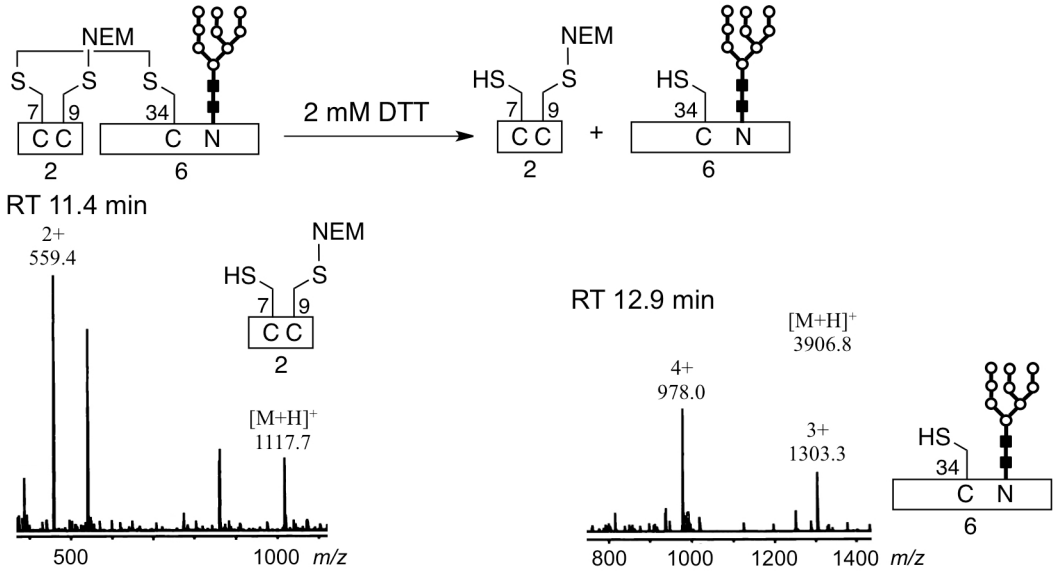


Figure S11. LC/MS spectra of fragments 2 and 6 of the peak c shown in Figure S8 after Lys-C digestion and DTT reduction.

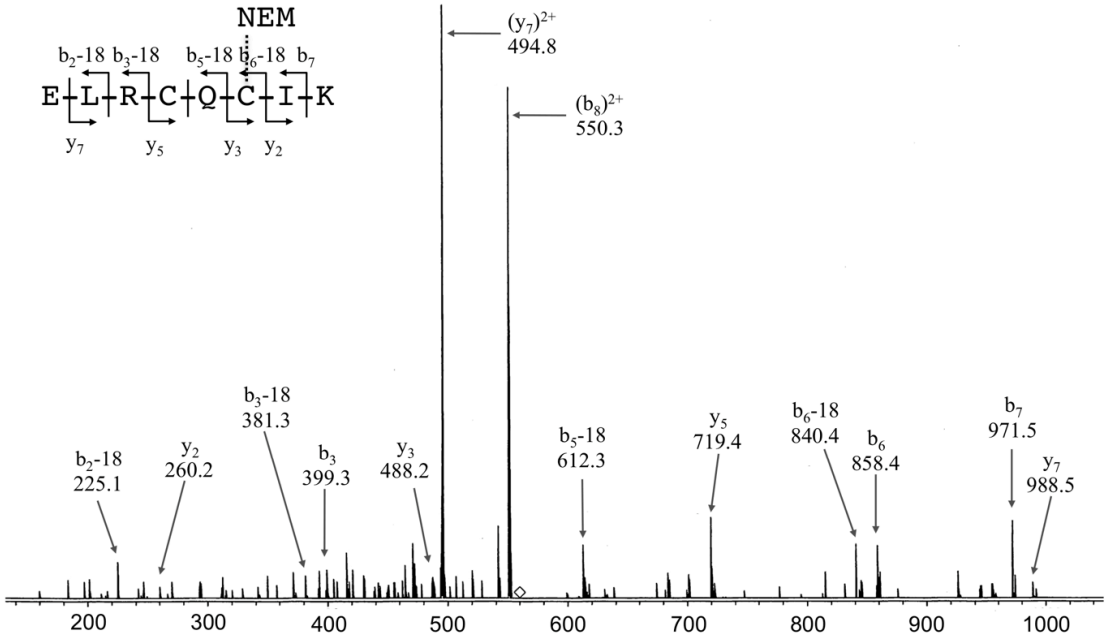


Figure S12. Tandem mass spectrum for the ion at *m/z* 559.4 (doubly charged) from the MS spectrum of RT 11.4 min of Figure S11 (amino acid 4–11). MS/MS analysis suggested that Cys-9 was alkylated with NEM.

Endopeptidase Lys-C digestion and DTT reduction of peak b of Figure S8.

Peak b of Figure S8 was also analyzed with the same procedures used for peak c shown in Figure S10-S12.

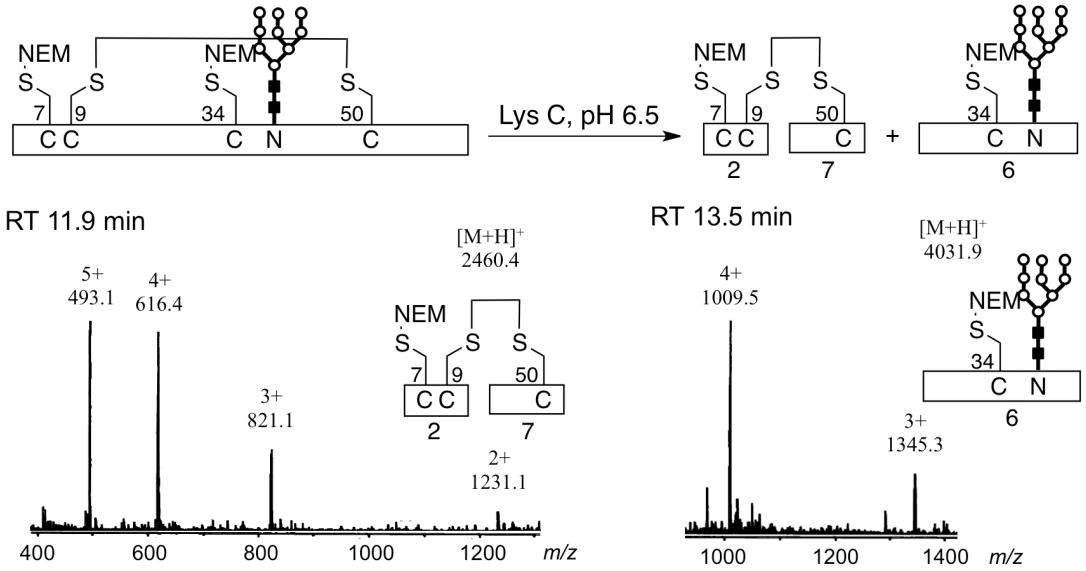


Figure S13. LC/MS spectra of cysteine-containing fragments (2, 6, 7) of the peak b shown in Figure S8 after Lys-C digestion.

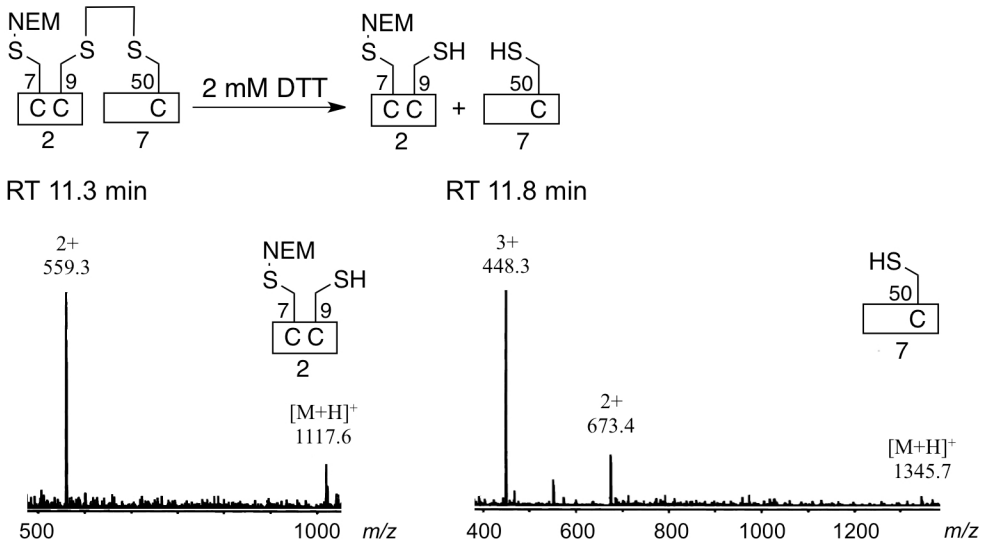


Figure S14. LC/MS spectra of fragments 2 and 7 of the peak b shown in Figure S8 after Lys-C digestion and DTT reduction.

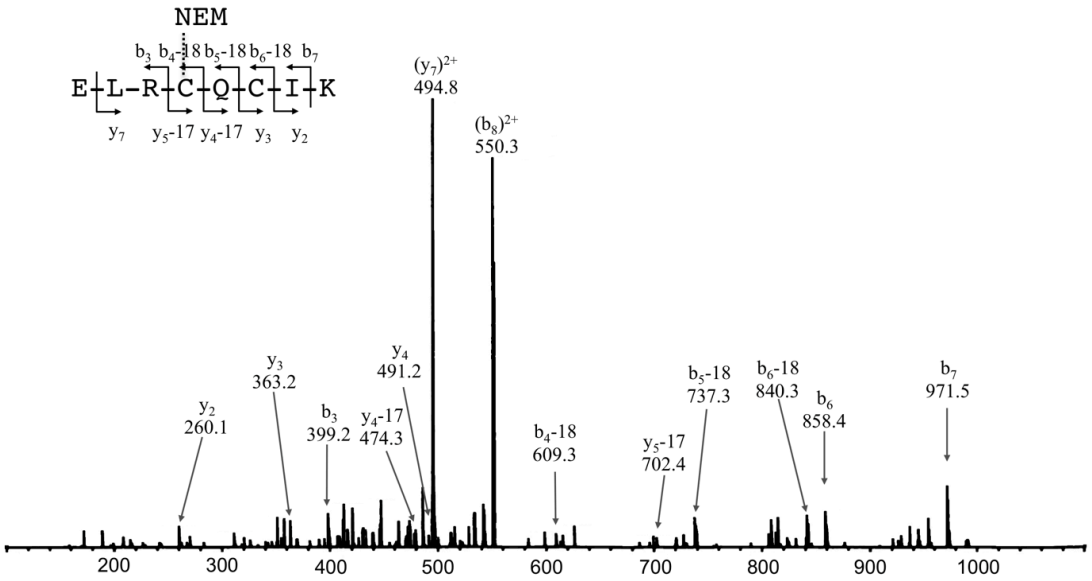


Figure S15. Tandem mass spectrum for the ion at *m/z* 559.4 (doubly charged) from the MS spectrum of RT 11.3 min of Figure S14 (amino acid 4–11). MS/MS analysis suggested that Cys-7 was alkylated with NEM.

Conclusion of disulfide bonds analysis of M9-IL-8 (6)

In the partially reduced peak c (Fig. S8 and S9), Cys9 in fragment 2 (Fig. S11 and S12) and Cys50 in fragment 7 (Fig. S10) were alkylated with NEM and Cys7 in fragment 2 (Fig. S11 and S12) and Cys34 in fragment 6 (Fig. S11) were not alkylated. In the partially reduced peak b (Fig. S8 and S9), Cys7 in fragment 2 (Fig. S14 and S15) and Cys34 in fragment 6 (Fig. S13) were alkylated with NEM and Cys9 in fragment 2 (Fig. S14 and S15) and Cys50 in fragment 7 (Fig. S14) were not alkylated. Therefore, disulfide bonds in M9-IL-8 **6** were determined to be Cys7–Cys 34 and Cys9–Cys50, which corresponds to the native linkages.

Preparation of misfolded M9-IL-8 glycoproteins. A solution of a folding buffer containing 1 M Gn•HCl, 0.1 M Tris–acetate, pH 8.3 was fleshly bubbled with air for 5 min. Glycopeptide **3** was dissolved in the folding buffer at the concentration of 0.2 mg/mL. In order to monitor this folding process, an aliquot at each defined time was diluted with 1/10 amount of 1 M HCl to stop the reaction and analyzed by HPLC (Vydac C4 Φ4.6 × 250 mm, 0.1% TFA : 0.1% TFA in 90% MeCN = 70 : 30 to 50 : 50 over 30 min at 1 mL/min) and LC/MS (Proteonavi Φ2.0 × 150 mm, 0.1% FA : 0.1% FA in 90% MeCN = 75 : 25 to 55 : 45 over 20 min at 0.2 mL/min). After 30 min, an aliquot was applied to HPLC using conditions mentioned above and folding intermediates **4** and **5** were isolated. After 4 h, the solution was directly applied to HPLC (Vydac C4 Φ4.6 × 250 mm, 0.04% TFA : 0.04% TFA in 90% MeCN = 75 : 25 for 10 min then 75 : 25 to 55 : 45 over 100 min at 1 mL/min) and fractions containing **6**, **7**, **8**, or **9** were collected and lyophilized.; ESI-MS: *m/z* calcd. for $C_{439}H_{717}N_{106}O_{161}S_4$ $[M+H]^+$ 10184.4, found 10183.8 (**4**), 10183.9 (**5**), *m/z* calcd. for $C_{439}H_{715}N_{106}O_{161}S_4$ $[M+H]^+$ 10182.4, found 10182.0 (**6**), 10181.8 (**7**), 10182.4 (**8**), *m/z* calcd. for **9** $C_{878}H_{1429}N_{212}O_{322}S_8$ $[M+H]^+$ 20363.8, found 20363.4 (deconvoluted).

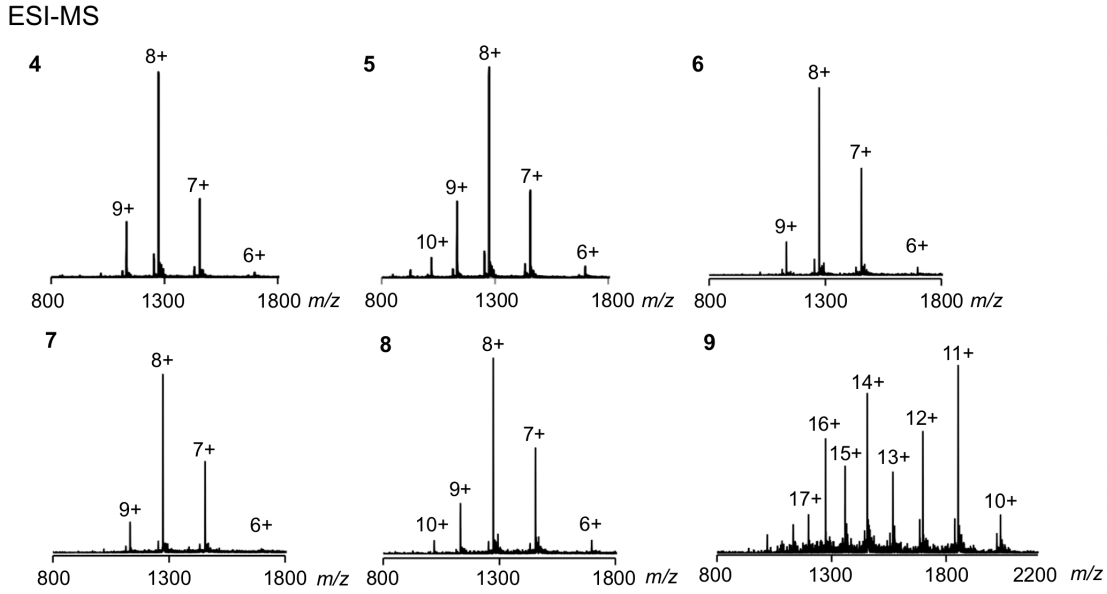


Figure S16. ESI-MS spectra of purified **4**, **5**, **6**, **7**, **8**, **9**.

Determination of disulfide linkages in the intermediates 4 and 5.

Intermediates **4** and **5** were confirmed to have a single disulfide bond individually based on mass analysis (Fig. S16). Without partial reduction by TCEP, intermediate **4** or **5** in 0.1 M Tris-HCl (pH 6.5) was treated with 10 mM NEM at 37ºC for 30 min, then endopeptidase Lys-C was added and the solution was kept at 37ºC for additional 30 min. Lys-C digestion gave 10 peptide fragments which were assigned from 1 to 10 and all fragments were detected by LC-MS (cadenza CD-C18 Φ2.0 × 100 mm, 0.1% FA : 0.1% FA in 90% MeCN = 95 : 5 to 40 : 60 over 20 min at 40ºC at 0.2 mL/min). Disulfide bond positions were determined by connection of cysteine-containing fragments (2, 6 and 7) as shown in Figure S17 and S18.

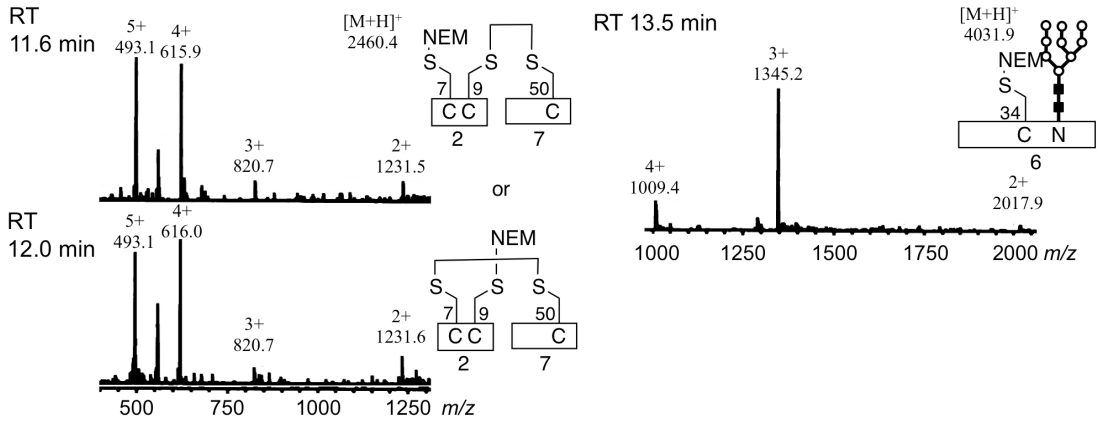


Figure S17. LC/MS spectra of cysteine-containing fragments (2, 6, 7) of NEM labeled **4** generated by Lys-C digestion. Since two LC peaks correspond to fragments 2 and 7 were detected at RT = 11.6 and 12.0 min, it was assumed that intermediate 4 is a mixture of Cys7-Cys50 isomer and Cys9-Cys50 isomer.

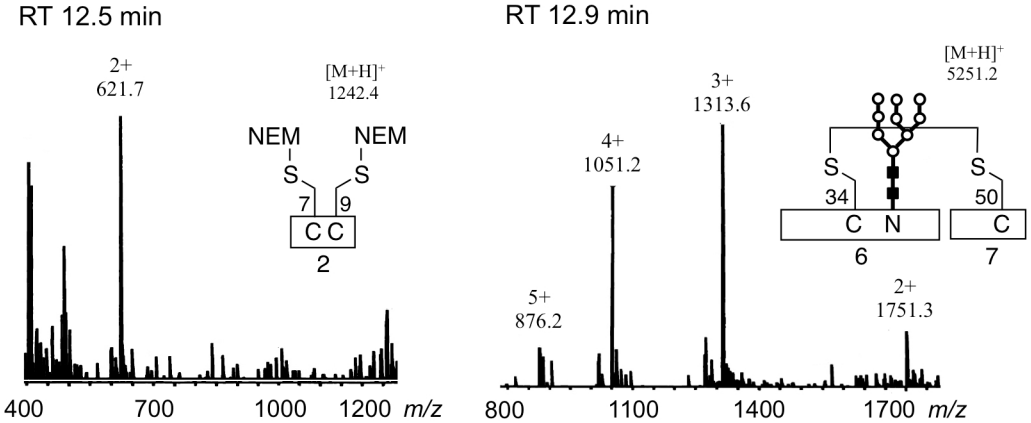


Figure S18. LC/MS spectra of cysteine-containing fragments (2, 6, 7) of NEM labeled **5** after digestion with Lys-C. Disulfide bond was found between Cys34 and Cys50.

Determination of disulfide linkages in misfolded M9-IL-8 (7)

M9-IL-8 (**7**) was also treated with endopeptidase Lys-C and fragments thus formed were analyzed by LC/MS.

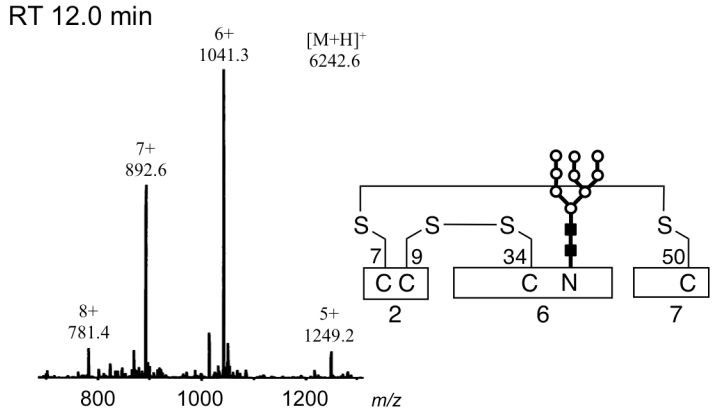


Figure S19. LC/MS spectrum of cysteine-containing fragments (2, 6, 7) of **7** after Lys-C digestion. MS spectrum suggested that three cysteine-containing fragments (2, 6, 7) were all linked together through disulfide bonds. Since the disulfide bond pairs of **6** were determined to be (Cys7–Cys34, Cys9–Cys50), those of **7** were determined to be (Cys7–Cys50, Cys9–Cys34).

Determination of disulfide linkages in misfolded M9-IL-8 (8)

M9-IL-8 **8** was also treated with endopeptidase Lys-C and fragments thus formed were analyzed by LC/MS.

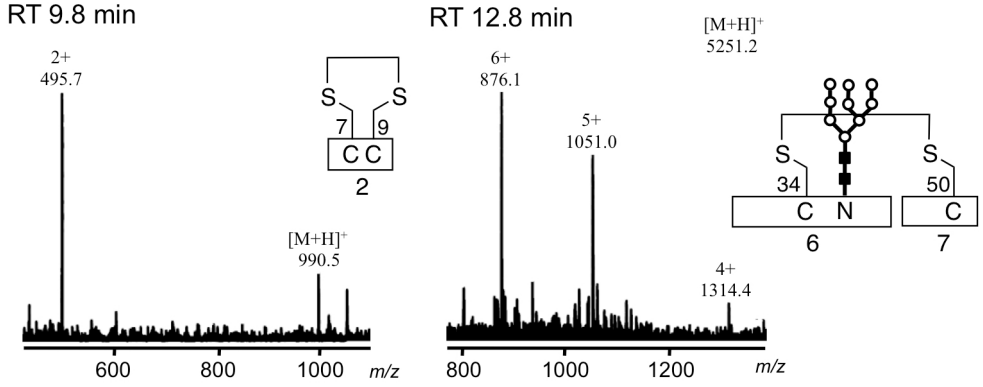


Figure S20. LC/MS spectra of cysteine-containing fragments (2, 6, 7) of **8** after Lys-C digestion. MS spectra suggested the formation of one disulfide bond within fragment 2 and another disulfide bond between fragments 6 and 7. Therefore, the disulfide bond pairs were determined to be (Cys7–Cys9, Cys34–Cys50).

Analysis of disulfide linkages in misfolded M9-IL-8 dimer (9)

M9-IL-8 dimer **9** was also treated with endopeptidase Lys-C and fragments thus formed were analyzed by LC-MS.

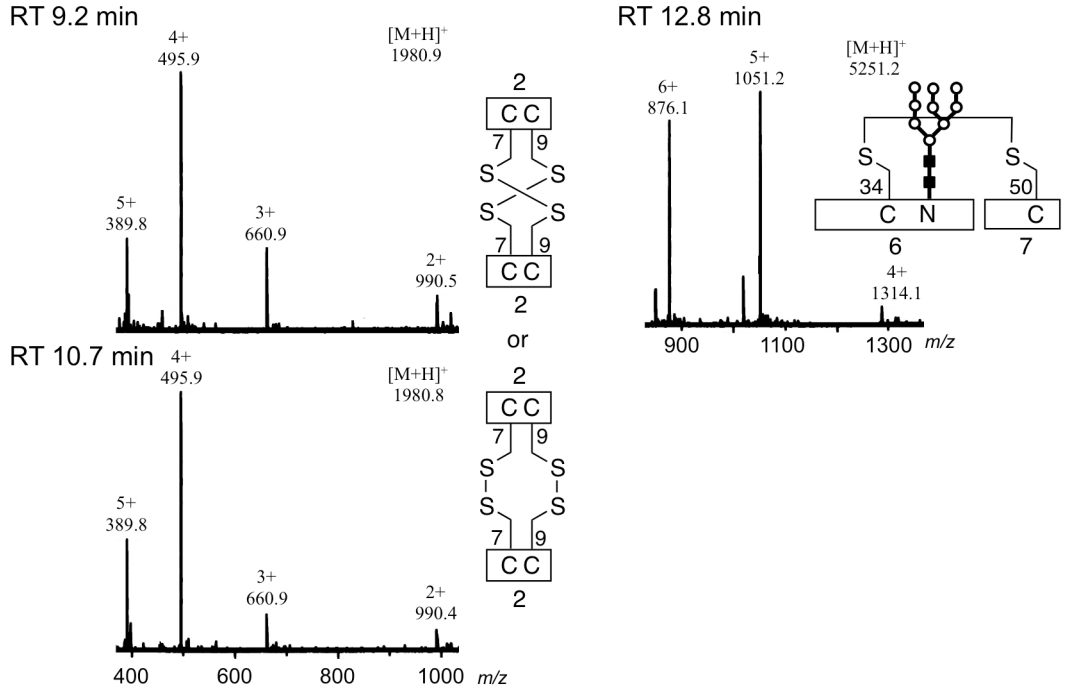


Figure S21. LC/MS spectra of cysteine-containing fragments (2, 6, 7) of **9** after Lys-C digestion. Two different fragment 2 dimers were observed at RT = 9.2 and 10.7 min together with fragments 6 and 7 connected via a disulfide bond. Therefore, it was assumed that M9-IL-8 peptides were dimerized at Cys7 and Cys9, and the intra-chain disulfide bond was formed between Cys34 and Cys50. It is probably a mixture of inter-chain disulfide bond pair isomers (Cys7-Cys7, Cys9-Cys9) and (Cys7-Cys9, Cys7-Cys9).

UGGT assay. UGGT reactions were carried out in 50 mM Tris-HCl, pH 7.5 containing 5 mM $CaCl_2$ and 0.5 mM UDP-Glc and 4 μM of glycoprotein substrate (**6**, **7**, or **9**) and 0.6 ng/μL UGGT and incubated at 37ºC. An aliquot was diluted with the same volume of 4% aqueous TFA to stop the reaction, and glucose transfer was analyzed by LC/MS (Proteonavi Φ2.0 × 150 mm, 0.1% FA : 0.1% FA in 90% MeCN = 75 : 25 to 55 : 45 over 20 min at 0.2 mL/min). The glucosyl transfer ratios were estimated from the abundance of the deconvoluted peak of the mass spectra. Starting material and glucosylated product were eluted at the same LC retention time. Substrate consumptions were estimated by deconvoluted MS peaks, assuming that M9- and G1M9 glycoprotein have identical ionization efficiency. Unfortunately, the other misfolded product **8** seemed to be not stable under UGGT assay conditions and the corresponding LC/MS peak was gradually diminished during the reaction.

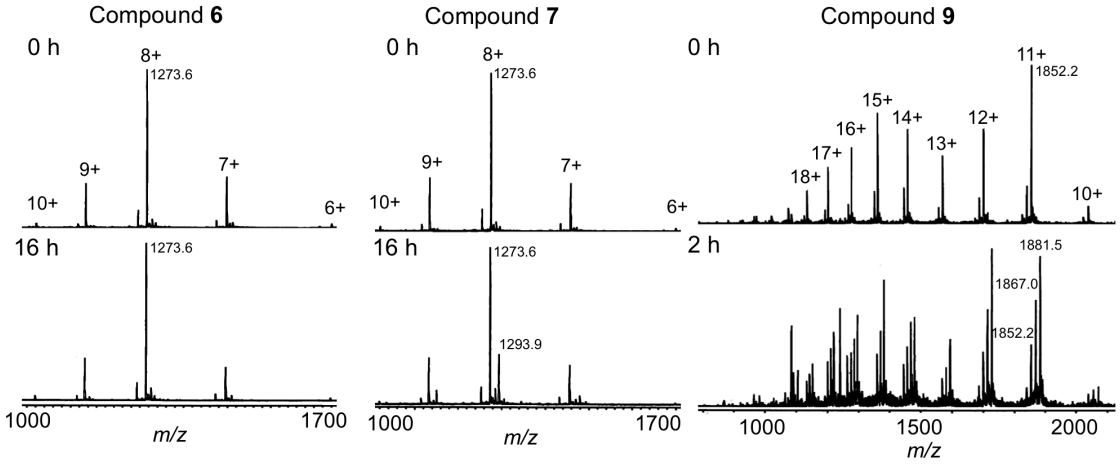


Figure S22. LC/MS spectra of UGGT reaction products of **6**, **7**, and **9**.

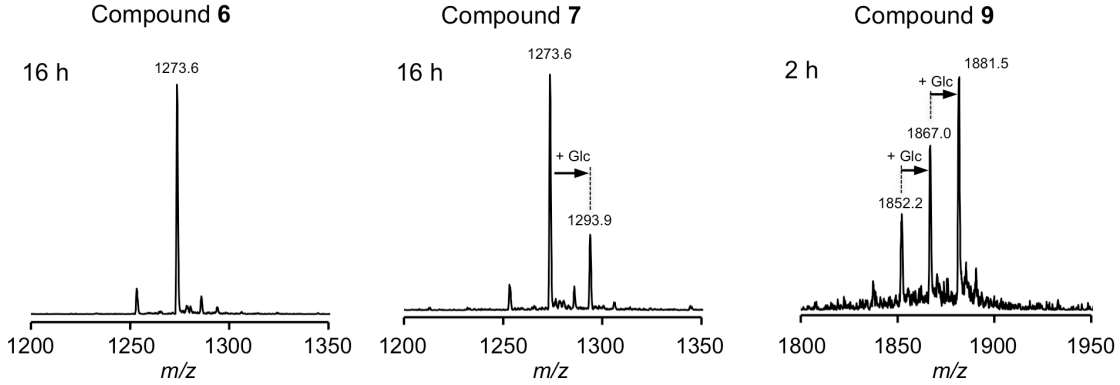


Figure S23. Expansion of LC/MS spectra of UGGT reaction products of **6**, **7**, and **9** from Figure S22.

References

(1) Semisotnov, G. V.; Rodionova, N. A.; Razgulyaev, O. I.; Uversky, V. N.; Gripas, A. F.; Gilmanshin, R. I. *Biopolymers* **1991**, *31*, 119.

(2) Kajihara, Y.; Suzuki, Y.; Yamamoto, N.; Sasaki, K.; Sakakibara, T.; Juneja, L. R. *Chem. Eur. J.* **2004**, *10*, 971.

(3) Seko, A.; Koketsu, M.; Nishizono, M.; Enoki, Y.; Ibrahim, H. R.; Juneja, L. R.; Kim, M.; Yamamoto, T. *Biochim. Biophys. Acta* **1997**, *1335*, 23.

(4) Koketsu, M.; Seko, A.; Juneja, L. R.; Kim, M.; Kashimura, N.; Yamamoto, T. *J. Carbohydr. Chem.* **1995**, *14*, 833.

(5) Sumiyoshi, W.; Nakakita, S.; Hasehira, K.; Miyanishi, N.; Kubo, Y.; Kita, T.; Hirabayashi, J. *Biosci. Biotechnol. Biochem.* **2010**, *74*, 606.

(6) Schnolzer, M.; Alewood, P.; Jones, A.; Alewood, D.; Kent, S. B. H. *Int. J. Pept. Res. Ther.* **2007**, *13*, 31.

(7) Hackeng, T. M.; Griffin, J. H.; Dawson, P. E. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 10068.

(8) Swerdloff, M. D.; Anderson, S. B.; Sedgwick, R. D.; Gabriel, M. K.; Brambilla, R. J.; Hindenlang, D. M.; Williams, J. I. *Int. J. Peptide Protein Res.* **1989**, *33*, 318.

(9) Yamamoto, N.; Takayanagi, Y.; Yoshino, A.; Sakakibara, T.; Kajihara, Y. *Chem. Eur. J.* **2007**, *13*, 613.

(10) Yamamoto, N.; Takayanagi, A.; Sakakibara, T.; Dawson, P. E.; Kajihara, Y. *Tetrahedron Lett.* **2006**, *47*, 1341.

(11) Bacsa, B.; Kappe, C. O. *Nat. Protoc.* **2007**, *2*, 2222.