Electronic Supplementary Material

Synthesis of misfolded glycoprotein dimers through native chemical ligation of dimeric peptide thioester

Masayuki Izumi*[a], Shinji Komaki[a], Ryo Okamoto[a], Akira Seko[b], Yoichi Takeda[b], Yukishige Ito*[bc], and Yasuhiro Kajihara*[ab]

[a]Department of Chemistry, Graduate School of Science, Osaka Univeristy, 1-1 Machikaneyama, Toyonaka, Osaka 560-0043, Japan. E-mail: mizumi@chem.sci.osaka-u.ac.jp / kajihara@chem.sci.osaka-u.ac.jp

[b]Japan Science and Technology Agency, ERATO Ito glycotrilogy project, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan. E-mail: yukito@riken.jp

[c]Synthetic Cellular Chemistry Laboratory, RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan.

Table of contents

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 Bachem AG. 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•HCl) were purchased from Tokyo Chemical Ind. N,N-dimethylformamide (DMF), dichloromethane (DCM), 4-mercaptophenylacetic acid (MPAA), *m*-cresol, thioanisole, trifluoromethanesulfonic acid (TFMSA), diethyl ether, D,L-dithiothreitol (DTT), endoprotease Lys-C (LysC), UDP-glucose 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-hydroxymethylphenoxyacetic acid (HMPA) were purchased from Novabiochem. N-Methyl-2-pyrrolidinone (NMP), Guanidine hydrochloride (Gn•HCl), and HPLC grade acetonitrile were purchased from Kanto Chemical. 2-Amino-2-hydroxymethyl-propane-1,3-diol (Tris) was purchased from Sigma-Aldrich Co. LLC.

LC-MS analysis, HPLC analysis and purification. LC-MS analyses were performed on a Bruker Daltonics amaZon-SL ion trap mass spectrometer equipped with an Agilent 1260 HPLC system, and UHPLC-MS analyses were performed on a Bruker Daltonics amaZon-ETD ion trap mass spectrometer equipped with a Thermo Scientific UltiMate 3000 UHPLC system. Gradient elution of water/0.1% formic acid (solvent A) versus 90% CH_3CN/0.1% formic acid (solvent B) was used. The eluent was monitored at 218 nm. For LC-MS and UHPLC-MS analyses, Cadenza CD-C18 (Imtakt, 2.0 × 75 mm) or Proteonavi (Shiseido, 2.0 × 150 mm or 2.0 × 50 mm) columns were used at the flow rate of 0.2 mL/min (LC-MS) or 0.3 mL/min (UHPLC-MS). RP-HPLC analyses were performed on a Waters 486 HPLC system equipped with 2996 photodiode array detector or Thermo Scientific Ultimate 3000 UHPLC system equipped with VWD-3400 variable wavelength detector. Gradient elution of water/0.1% TFA (solvent A) versus 90% CH_3CN/0.1% TFA (solvent B) was used and the eluent was monitored at 220 nm or 216 nm. Cadenza CD-C18 (Imtakt, 4.6 × 75 mm) or Proteonavi (Shiseido, 4.6 × 250 mm) columns were used at the flow rate of 1.0 mL/min. For semipreparative HPLC, Vydac C18 (Grace Vydac, 10 × 250 mm) or Proteonavi (Shiseido, 10 × 250 mm) columns were used at the flow rate of 2.0 mL/min. ESI-MS analyses of purified peptides were performed on a Bruker Daltonics Esquire 3000 plus ion trap mass spectrometer.

Amino acid sequence of M9-IL-8

```
          10         20         30        40
SAKELRCQCI KTYSKPFHPK FIKELRVIES GPACANTEII

          50         60         70 72
VKLSDGRELC LDPKENWVQR VVEKFLKRAE NS
```

M9-IL-8 was prepared from two segments IL-8(S1-A33) and M9-IL-8(C34-S72): AC typed in boldface is a ligation junction. Cys7 (underlined) in the native sequence of IL-8 was mutated to Ser to simplify the crosslinking by bismaleimide linker for the synthesis of M9-IL-8 dimers. Asn36 (underlined) was glycosylated with M9 oligosaccharide.

Synthesis of IL-8(1-33) peptide thioester 2. IL-8(1-33) peptide thioester was synthesized according to the published procedure[1] on 3-mercaptopropionamide-Leu-OCH_2-Pam resin using in situ neutralization Boc solid phase peptide synthesis (SPPS).

Synthesis of 7,9-bis(*S*-carbamidomethyl)-IL-8(1-33) peptide thioester 3. IL-8(1-33) peptide thioester **2** (0.5 mg, 0.12 μmol) was dissolved in a buffer containing 6 M Gn•HCl, 0.2 M Na_2HPO_4, 5 mM iodoacetamide, pH 7.3 (0.15 mL). The solution was stand for 1.5 h at r.t. The solution was diluted with 0.04% aqueous TFA and purified by semipreparative HPLC (Proteonavi, solvent A:solvent B =85:15 to 65:35 over 100 min at 2 mL/min) and thioester **3** (0.5 mg, quant) was obtained after lyophilization.; ESI-MS *m/z* calcd. for $C_{178}H_{290}N_{49}O_{50}S_4$ $[M + H]^+$ 4044.8, found 4043.1.

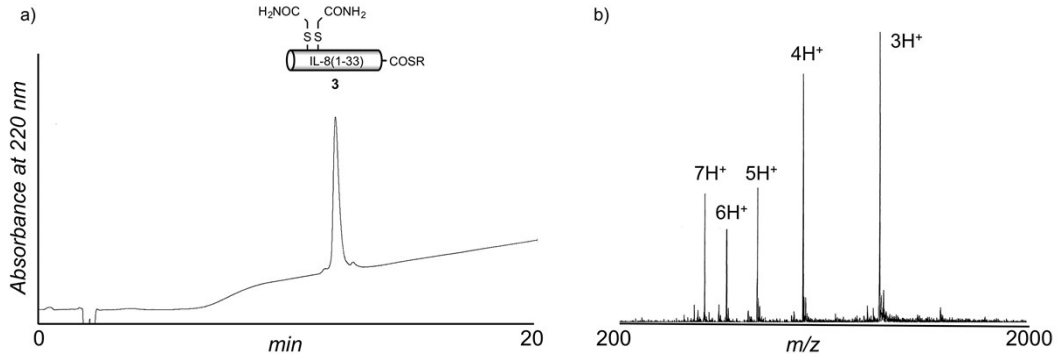


Fig. S1. Analytical data of the preparation of **3**. a) RP-HPLC profile of purified **3**. b) ESI-MS spectrum of purified **3**.

Synthesis of M9-IL-8(34-72) glycopeptide 4. Glycopeptide **4** bearing M9 oligosaccharide was synthesized according to the published procedure[1] on HMPA-PEGA resin using Fmoc-Asn(M9 oligosaccharide)-OH prepared from hen egg yolk.

Synthesis of 7,9-bis(*S*-carbamidomethyl)-M9-IL-8(34-50:disulfide) 5. 7,9-bis(*S*-carbamidomethyl)-IL-8(1-33) thioester **3** (0.7 mg, 0.17 μmol) and M9-IL-8(34-72) **4** (0.9 mg, 0.14 μmol) was dissolved in a freshly degassed buffer containing 6 M Gn•HCl, 0.2 M Na_2HPO_4, 0.1 M MPAA, 20 mM TCEP (pH 6.6) (87 μL). After being stand overnight, a 10% solution of DTT in water (8 μL) was added to the solution. The solution was directly applied to semipreparative HPLC (Proteonavi, solvent A:solvent B = 75:25 to 55: 45 over 100 min at 2 mL/min) to yield ligated polypeptide (~0.9 mg) after lyophilization. The ligated polypeptide was dissolved in a buffer containing 1 M Gn•HCl, 0.1 M Tris-acetate (pH 8.5) (4 mL) that was freshly bubbled with air for 5 min. After 9 h, the solution was directly injected to semipreparative HPLC (Proteonavi, solvent A:solvent B = 75:25 to 55: 45 over 100 min) to yield **5** (0.7 mg, 46%) after lyophilization.; ESI-MS *m/z* calcd for $C_{443}H_{723}N_{108}O_{163}S_4$ $[M + H]^+$ 10298.4, found 10299.1.

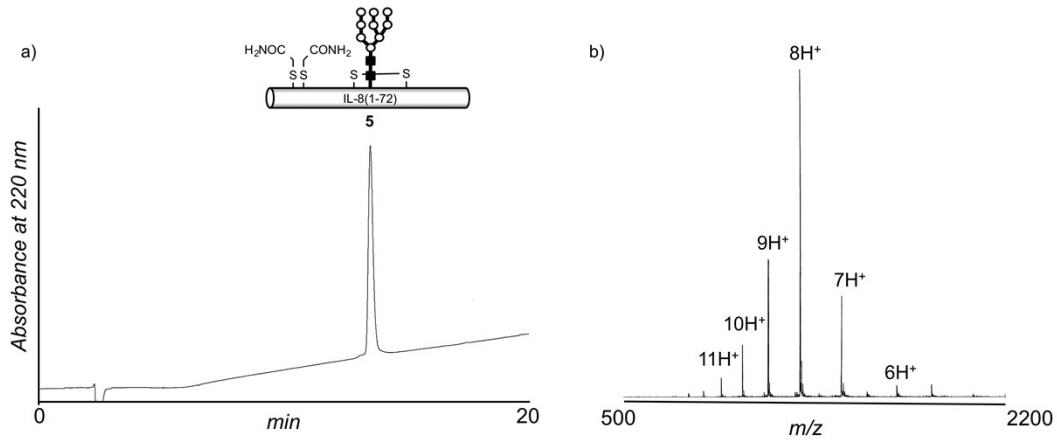


Fig. S2. Analytical data of the preparation of **5**. a) RP-HPLC profile of purified **5**. b) ESI-MS spectrum of purified **5**.

Synthesis of 7,9,34,50-tetrakis(*S*-carbamidomethyl)-M9-IL-8 7. Full-length M9-IL-8 polypeptide **6** (~0.1 mg) was dissolved in a buffer containing 1 M Gn•HCl, 0.1 M Tris-acetate (pH 8.5) (0.5 mL). After 30 min, 100 mM iodoacetamide in water (25 μL) and aqueous ammonia (10 μL) was added to the solution. After being stand for 1 h at r.t., the solution was purified by HPLC (Vydac C4, solvent A:solvent B = 75:25 to 55: 45 over 60 min at 1 mL/min) twice to yield **7** after lyophilization.; ESI-MS *m/z* calcd for $C_{447}H_{731}N_{110}O_{165}S_4$ $[M + H]^+$ 10414.5, found 10414.5.

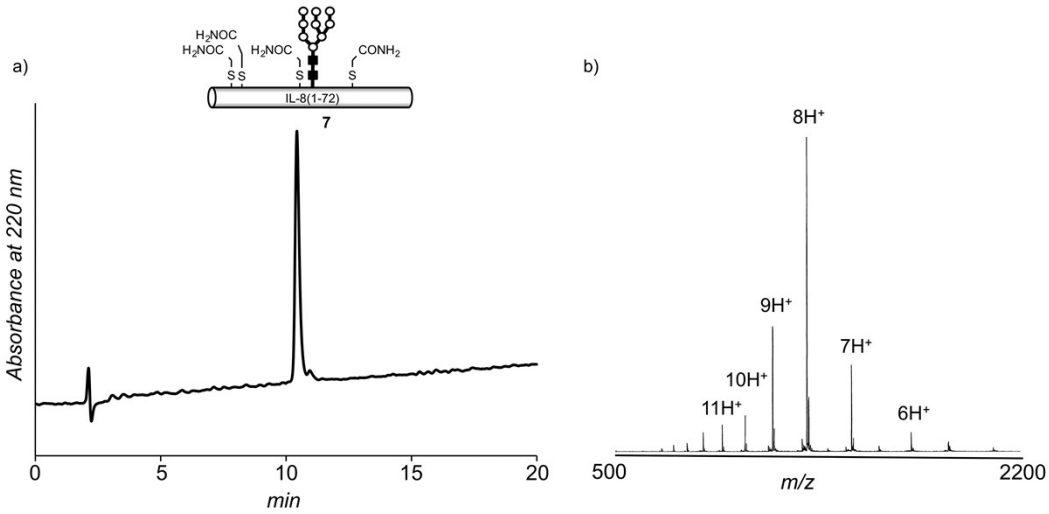


Fig. S3. Analytical data of the preparation of **7**. a) RP-HPLC profile of purified **7**. b) ESI-MS spectrum of purified **7**.

Synthesis of M9-IL-8(34–72, 34:50:disulfide) 8. M9-IL-8(34–72) glycopeptide **4** was dissolved in a buffer containing 1 M Gn•HCl, 0.1 M Tris-acetate (pH 8.5) that was freshly bubbled with air for 5 min. The concentration of the solution was adjusted to ~0.2 mg protein/mL. After 3.5 h, the solution was directly injected to semipreparative HPLC (Proteonavi, solvent A:solvent B = 75:25 to 55: 45 over 100 min at 2 mL/min) to yield **8** after lyophilization.; ESI-MS *m/z* calcd for $C_{267}H_{440}N_{59}O_{116}S_2$ $[M + H]^+$ 6396.8, found 6396.9.

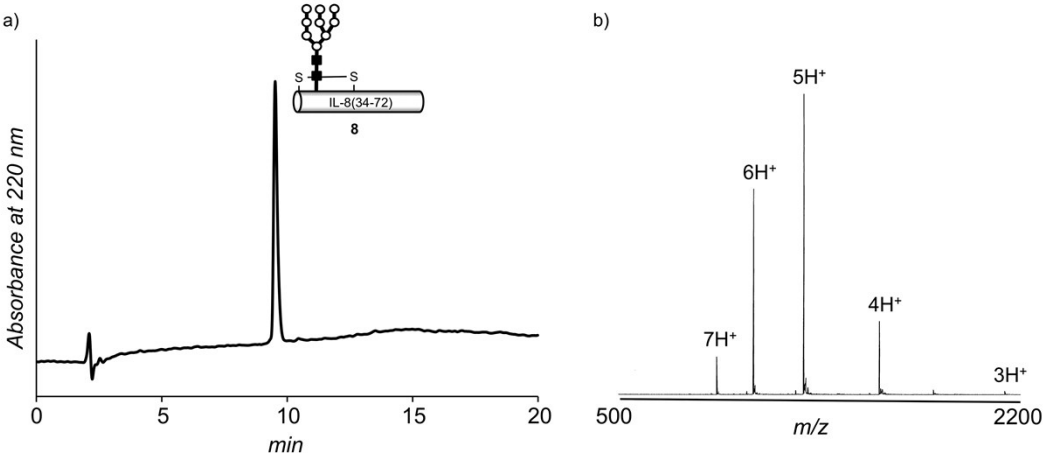


Fig. S4. Analytical data of the preparation of **8**. a) RP-HPLC profile of purified **8**. b) ESI-MS spectrum of purified **8**.

Synthesis of 34,50-bis(*S*-carbamidomethyl)-M9-IL-8(34-72) 9. M9-IL-8(34–72) peptide **4** (0.6 mg) was dissolved in a buffer containing 1 M Gn•HCl, 0.1 M Tris-acetate, pH 8.5, 5 mM iodoacetamide (0.6 mL). The solution was stand for 1 h at r.t. The solution was diluted with 4% aqueous TFA and purified by HPLC (Vydac C4, solvent A:solvent B =73:27 to 53:47 over 60 min at 1 mL/min) and **9** was obtained after lyophilization.; ESI-MS *m/z* calcd. for $C_{271}H_{448}N_{61}O_{118}S_2$ $[M + H]^+$ 6512.9, found 6512.9.

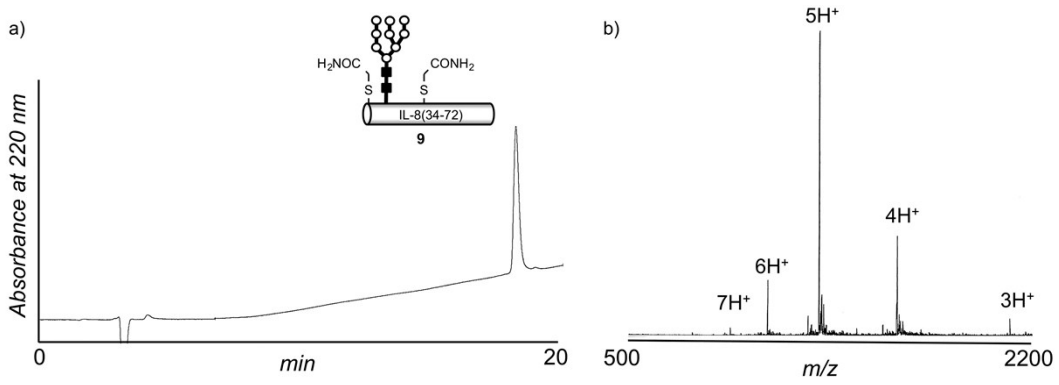
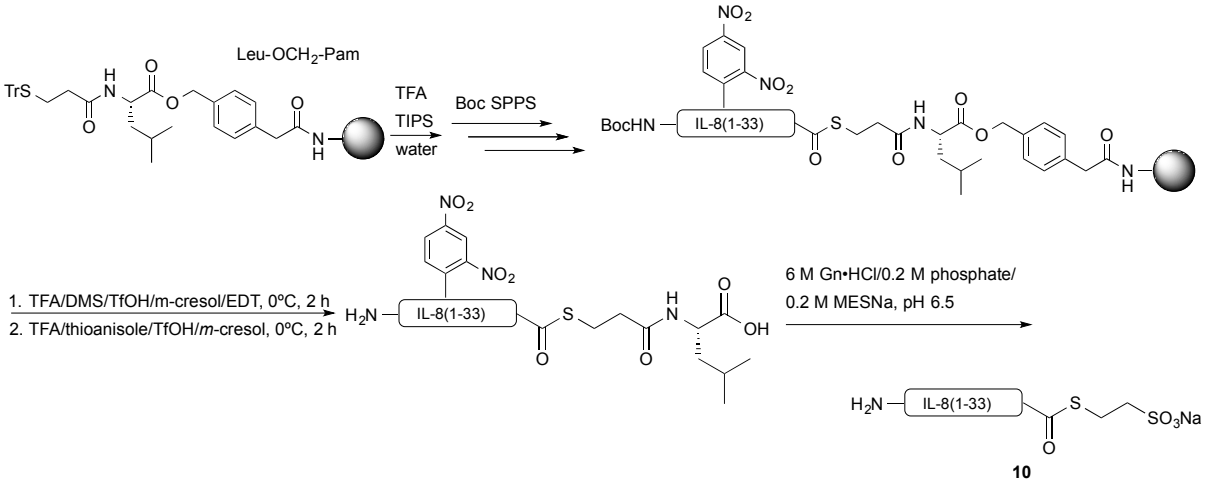


Fig. S5. Analytical data of the preparation of **9**. a) RP-HPLC profile of purified **9**. b) ESI-MS spectrum of purified **9**.

Synthesis of IL-8(1-33, C7S) MESNa thioester (10).



IL-8 (1-33, C7S) MESNa thioester **10** was prepared using the same reported procedure for the synthesis of IL-8 (1-33) MESNa thioester **2**.[1] Briefly, SPPS of the peptide thioester was carried out on Boc-Leu-OCH_2-PAM resin. After deprotection of the Boc group by treatment with neat TFA (1 min x 2), the resin was coupled with trityl 3-mercaptopropionic acid using HBTU and DIEA in DMF. Trityl group was removed by treatment with TFA/water/TIPS = 95/2.5/2.5 (1 min x 2), and the first Boc-amino acid was coupled using HBTU and DIEA in DMF for 1 h. Further peptide elongation was carried out using in situ

neutralization Boc protocol. After all amino acid coupling, the resin was treated with TFA/DMS/*m*-cresol/TfOH/EDT = 5/3/1/1/0.2 for 2 h at 0 ºC and then with TFA/m-cresol/thioanisole/TfOH = 10/1/1/1 for 2 h at 0 ºC. Crude peptide was precipitated from the cocktail with cold ether. The precipitate was purified by semipreparative HPLC and then treated with 0.2 M MESNa in 6 M Gn•HCl, 0.2 M Na_2HPO_4 at pH 6.5. The desired peptide thioester **10** (15.5 mg, 4%) was obtained after purification by semipreparative HPLC (Vydac C18, solvent A:solvent B =90:10 to 80:10 over 10 min then to 55: 45 over 120 min) and lyophilization.; ESI-MS *m/z* calcd. for $C_{174}H_{284}N_{47}O_{49}S_3$ $[M + H]^+$ 3912.0, found 3911.6.

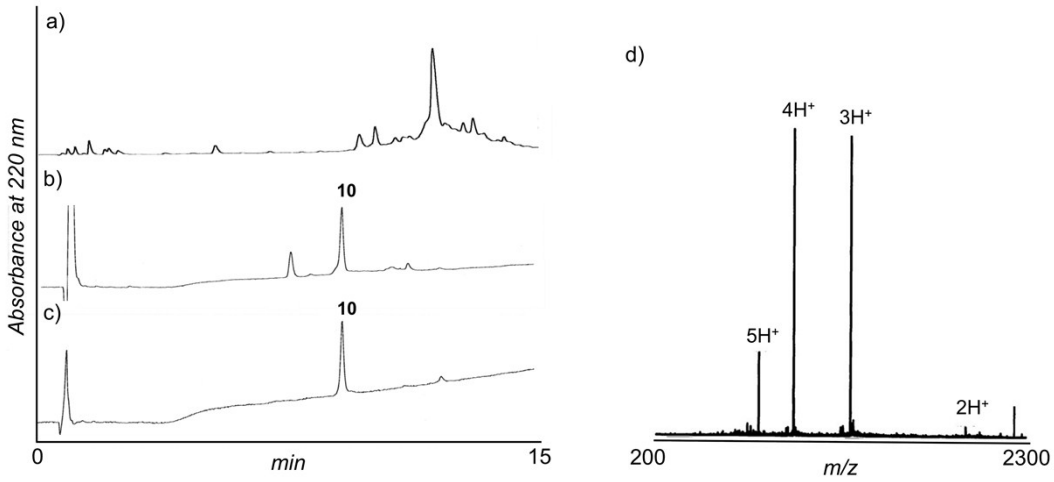


Fig. S6. Analytical data of the preparation of **10**. RP-HPLC profiles of a) crude precipitate after deprotection–cleavage of the peptide from a PAM resin, b) treatment of purified peptide with MESNa for the removal of a 2,4-dinitrophenyl group from a His residue. c) purified **10**. d) ESI-MS spectrum of purified **10**.

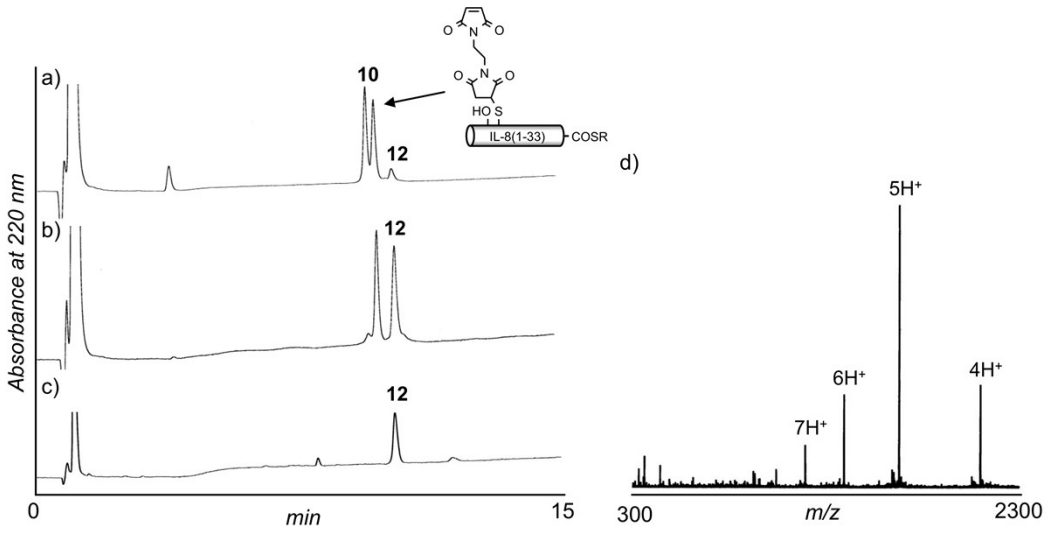


Fig. S7. Analytical data of the synthesis of dimeric bismaleimide-linked IL-8(1-33, C7S) thioester **12**. RP-HPLC profiles of the reaction, a) after 1 min, the peak between **10** and **12** corresponds to a peptide **10** reacted with bismaleimide linker (structure in the inset), b) after 1 h, c) purified **12**, d) ESI-MS spectrum of purified **12**.

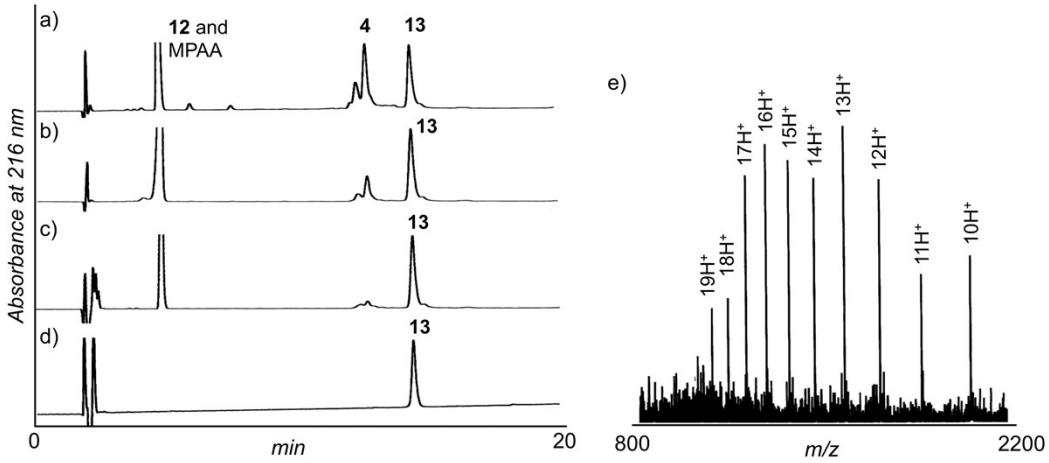


Fig. S8. Analytical data of the dNCL of IL-8(1-33) thioester dimer **12** to M9-IL-8(34-72) glycopeptide **4**. RP-HPLC profiles of the reaction, a) after 1 min, b) after 2.5 h, c) after 18 h and DTT treatment, d) purified **13**. e) ESI-MS spectrum of purified **13**.

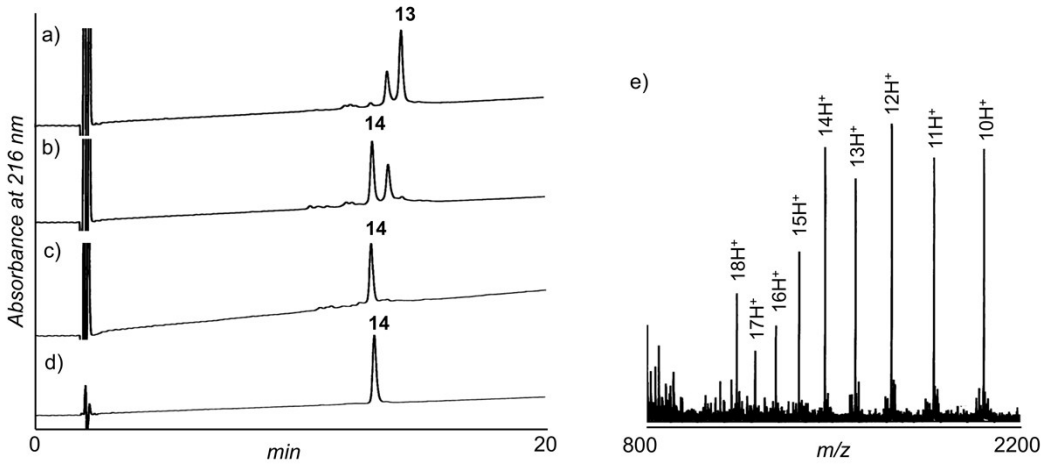


Fig. S9. Analytical data of disulfide formation of bismaleimide-linked homodimer having two M9 oligosaccharides **13**. RP-HPLC profiles of the reaction, a) after 1min, the peak on the left of **13** corresponds to a mono disulfide derivative, b) after 2.5 h, c) after 46 h, d) purified **14**. e) ESI-MS spectrum of purified **14**.

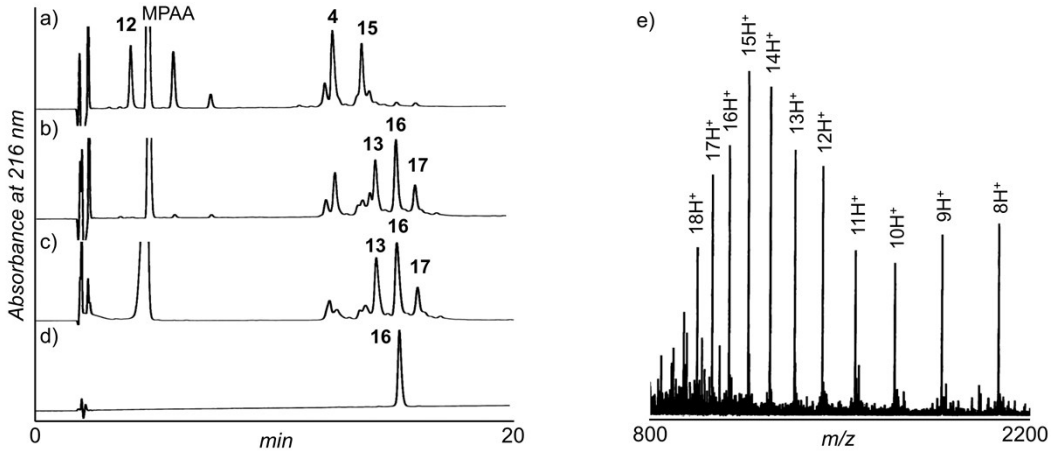


Fig. S10. Analytical data of the dNCL of the thioester dimer **12** to M9-IL-8(34-72) glycopeptide **4** and IL-8(34-72) peptide **15**. RP-HPLC profiles of the reaction, a) after 1 min, b) after 2.5 h, c) after 17 h and DTT treatment, d) purified heterodimer **16**. e) ESI-MS spectrum of purified **16**.

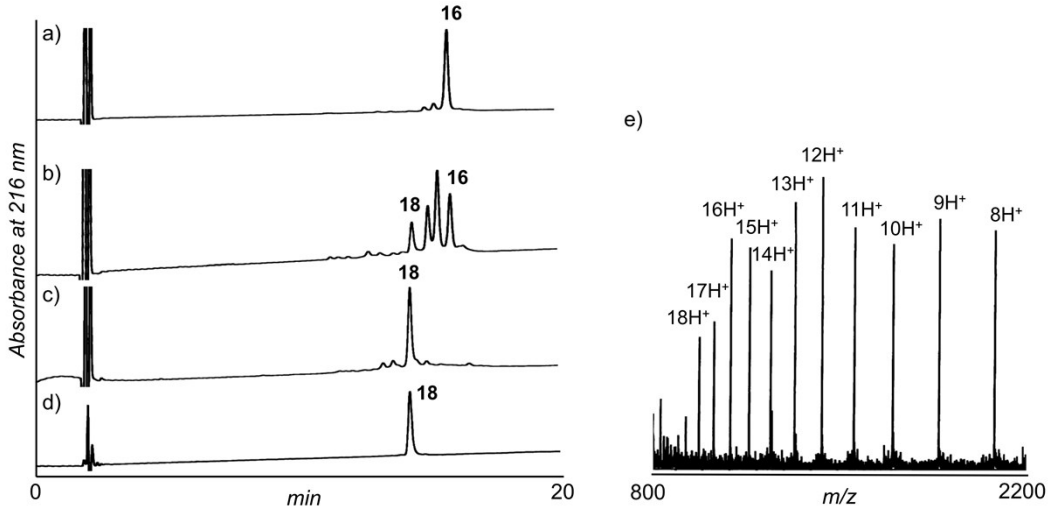


Fig. S11. Analytical data of the disulfide formation of the bismaleimide-linked heterodimer having one M9 oligosaccharide **16**. HPLC profiles of the reaction, a) after 1 min, b) after 2.5 h, two peaks between **18** and **16** correspond to mono disulfide derivatives, c) after 46.5 h, d) purified **18**. e) ESI-MS spectrum of purified **18**.

Disulfide mapping of homodimer 14 and heterodimer 18. Homodimer **14** (about 1 μg) was dissolved in 0.1 M Tris-HCl, pH 6.4 (19 μL) and LysC (0.1 μg/μL, 1 μL) was added. The solution was incubated for 4 h at 37 ºC and an aliquot was analyzed by LC-MS (Cadenza CD-C18, solvent A:solvent B = 98:2 to 35:65 over 17 min). Digestion of heterodimer **18** and LC-MS analysis was also carried out as the same way.; ESI-MS of the fragment of homodimer **14**; *m/z* calcd. for $C_{214}H_{363}N_{42}O_{104}S_2$ $[M + H]^+$ 5252.5, found 5251.5. the fragment of heterodimer **18** with M9 oligosaccharide; *m/z* calcd. for $C_{214}H_{363}N_{42}O_{104}S_2$ $[M + H]^+$ 5252.5, found 5252.4, without oligosaccharide *m/z* calcd for $C_{144}H_{247}N_{40}O_{49}S_2$ $[M + H]^+$ 3384.8, found 3385.0, the fragment containing bismaleimide linker of **14** and **18**; *m/z* calcd. for $C_{90}H_{159}N_{28}O_{32}S_2$ $[M + H]^+$ 2208.1, found 2208.2.

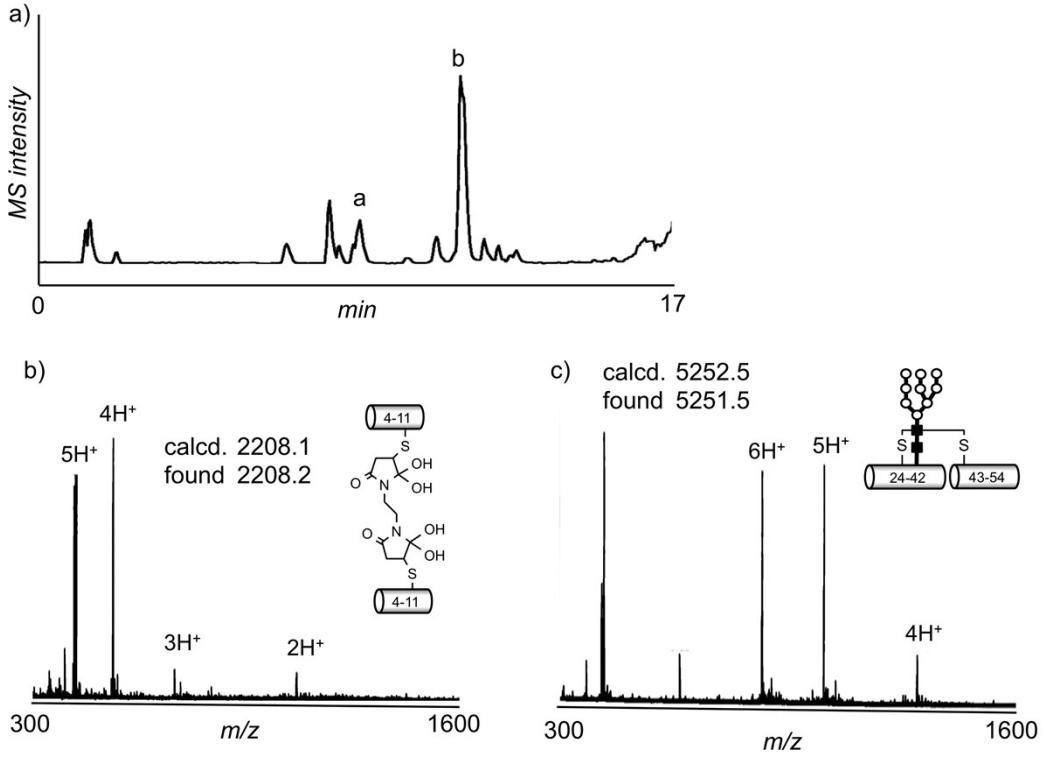


Fig. S12. a) LC-MS base peak chromatogram of the fragments of homodimer **14** after endoprotease LysC digestion. b) ESI-MS spectra of b) peak a, c) peak b. (insets are schematic representation of each fragment structure.)

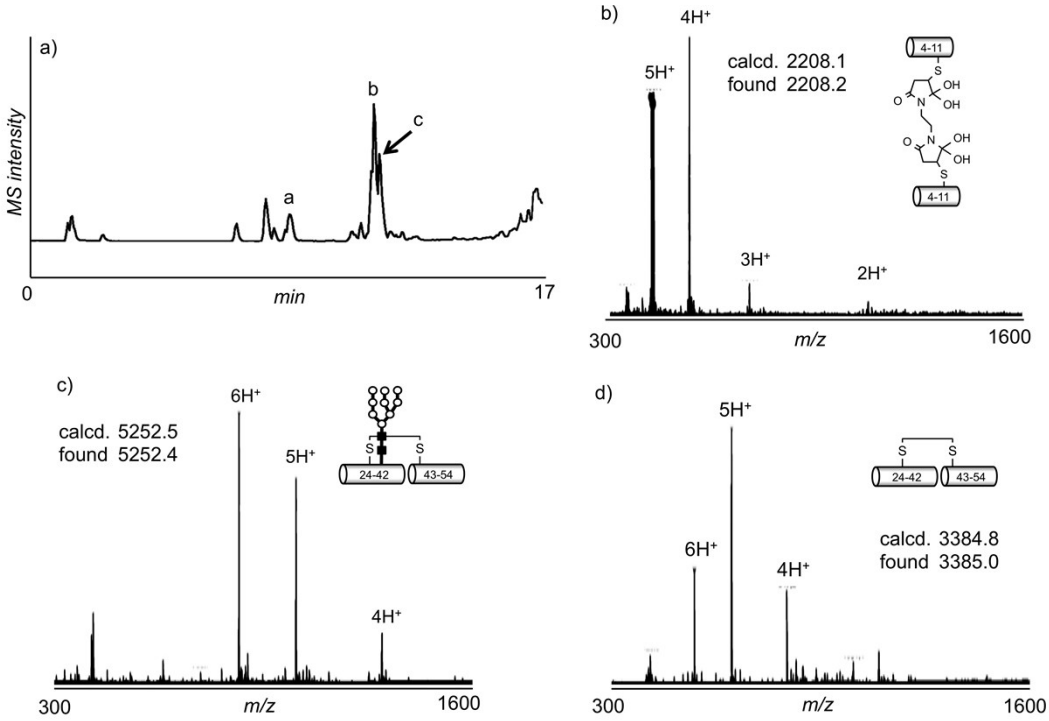


Fig. S13. a) LC-MS base peak chromatogram of the fragments of the heterodimer **18** after LysC digestion. b) ESI-MS spectra of b) peak a, c) peak b, and d) peak c. (insets are schematic representation of each fragment structure.)

CD spectra. Far-UV CD spectra were measured with a JASCO-J820 CD spectropolarimeter using 1 mm cuvette. Concentrations of glycoproteins were 5 μM or 2.5 μM. All measurements were carried out in 10 mM Tris-HCl, 5 M $CaCl_2$, pH 7.5.

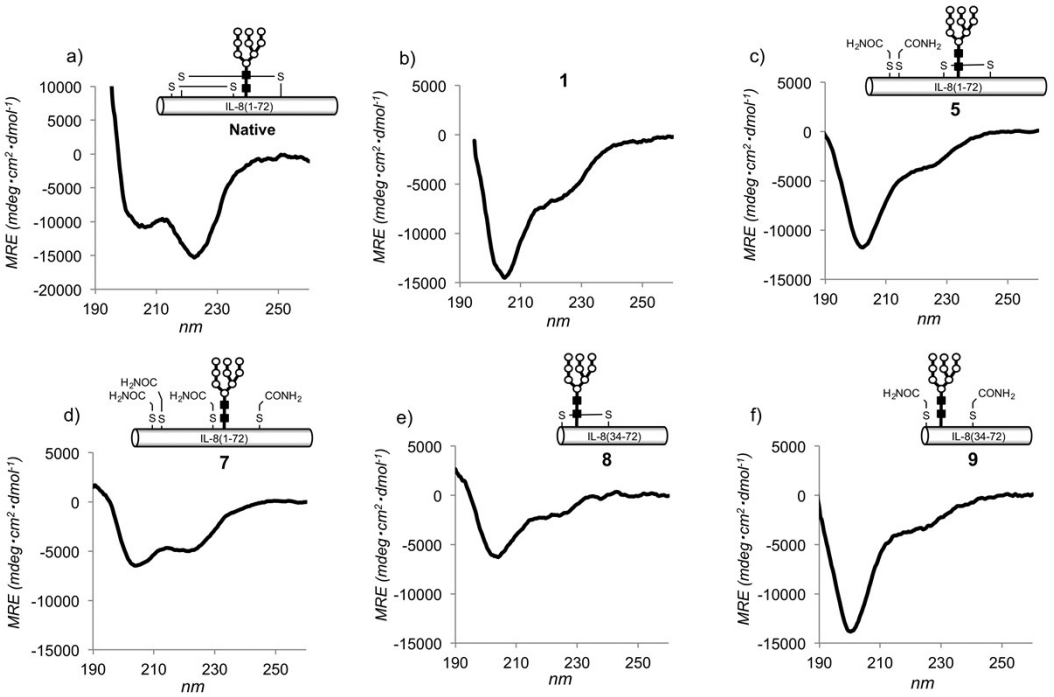


Fig. S14. CD spectra of a) native M9-IL-8, b) disulfide linked M9-IL-8 dimer **1**, c) **5**, d) **7,** e) **8** and f) **9**. (insets are schematic representation of each compounds.)

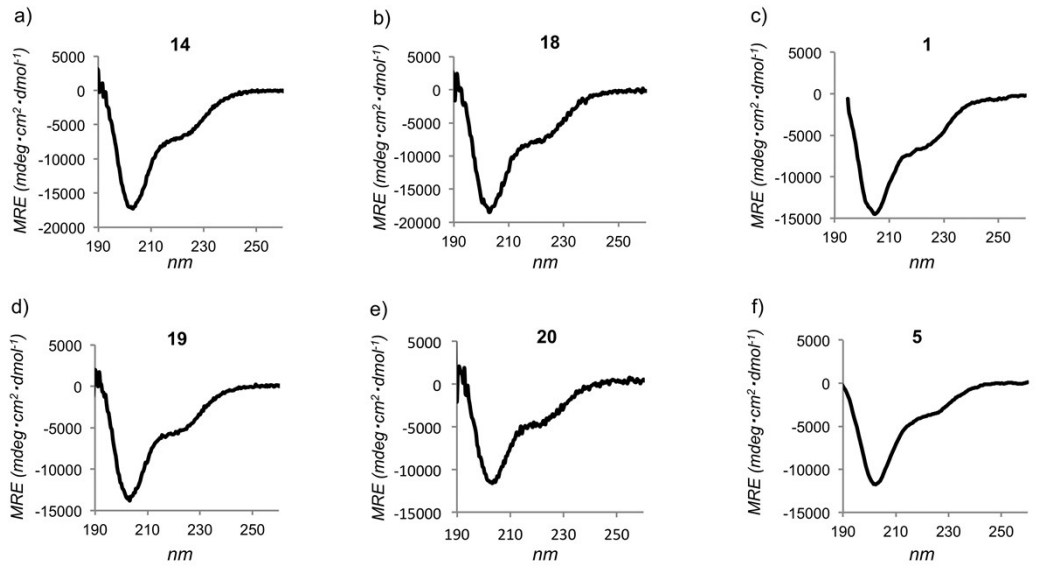


Fig. S15. CD spectra of a) bismaleimide–linked M9-IL-8 homodimer **14** and b) heterodimer **18**, c) disulfide linked M9-IL-8 dimer **1**, d) bismaleimide–linked G1M9-IL-8 homodimer **19** and e) heterodimer **20**, f) monomer **5**.

UGGT assay for kinetic analysis of *A. oryzae* UGGT.

UGGT reactions were performed in 50 mM Tris-HCl (pH 7.5), 5 mM $CaCl_2$, 0.5 mM UDP-glucose at 37 ºC for 15 min. Concentration of *A. oryzae* UGGT was adjusted for each substrate so that glucose transfer yield was less than 15% after 15 min reaction. The reaction was stopped by adding 0.1% TFA. An aliquot was analyzed by LC-MS (Proteonavi (2.0 × 50 mm), solvent A:solvent B = 71:29 to 56:44 over 5 min). Glucose transfer yield was estimated from the intensities of the peak of M9 and G1M9 derivatives from the MS spectrum. Each assay was triplicated.

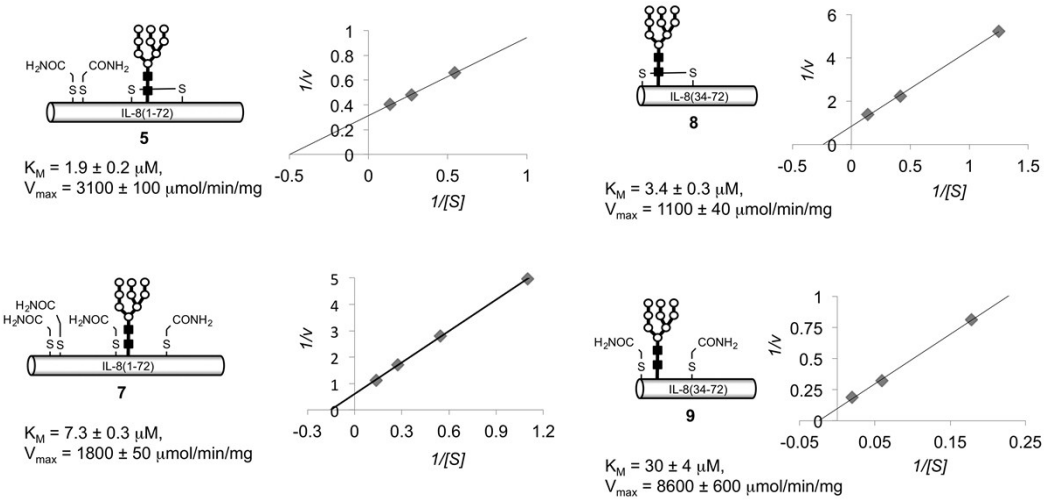


Fig. S16. Lineweaver-Burk plots of **5**, **7**–**9** by *A. oryzae* UGGT.

Reference

1. M. Izumi, Y. Makimura, S. Dedola, A. Seko, A. Kanamori, M. Sakono, Y. Ito and Y. Kajihara, *J. Am. Chem. Soc.*, 2012, **134**, 7238-7241.