DOI: 10.1021/bi902100f

BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

Correct Disulfide Pairing Is Required for the Biological Activity of Crustacean Androgenic Gland Hormone (AGH): Synthetic Studies of AGH[†]

Hidekazu Katayama,*[,‡] Hironobu Hojo,*[,‡] Tsuyoshi Ohira,[§] Akira Ishii,[‡] Takamichi Nozaki,[‡] Kiyomi Goto,[‡] Yuko Nakahara,[‡] Tetsuo Takahashi,[‡] Yuriko Hasegawa,[∥] Hiromichi Nagasawa,[⊥] and Yoshiaki Nakahara[‡]

[‡]*Department of Applied Biochemistry, Institute of Glycoscience, Tokai University, 1117 Kitakaname, Hiratsuka, Kanagawa 259-1292, Japan,* [§]*Department of Biological Sciences, Faculty of Science, Kanagawa University, Hiratsuka, Kanagawa 259-1293, Japan,* [∥]*Department of Biology, Keio University, Yokohama, Kanagawa 223-8521, Japan, and* [⊥]*Department of Applied Biological Chemistry, Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo, Tokyo 113-8657, Japan*

Received December 7, 2009; Revised Manuscript Received January 20, 2010

ABSTRACT: Androgenic gland hormone (AGH) of the woodlouse, *Armadillidium vulgare*, is a heterodimeric glycopeptide. In this study, we synthesized AGH with a homogeneous N-linked glycan using the expressed protein ligation method. Unexpectedly, disulfide bridge arrangement of a semisynthetic peptide differed from that of a recombinant peptide prepared in a baculovirus expression system, and the semisynthetic peptide showed no biological activity *in vivo*. To confirm that the loss of biological activity resulted from disulfide bond isomerization, AGH with a GlcNAc moiety was chemically synthesized by the selective disulfide formation. This synthetic AGH showed biological activity *in vivo*. These results indicate that the native conformation of AGH is not the most thermodynamically stable form, and correct disulfide linkages are important for conferring AGH activity.

Sex differentiation in crustaceans was first shown to be under hormonal control by Charniaux-Cotton in 1954 (*1*). This intrinsic factor, termed androgenic gland hormone (AGH), is synthesized in and secreted from the male-specific organ, the androgenic gland. In 1999, the chemical structure of AGH from a terrestrial isopod, *Armadillidiyum vulgare*, was determined to be a heterodimeric glycopeptide with an insulin-like amino acid sequence (*2*), and a cDNA encoding the AGH precursor was cloned (*3*). Up to now, cDNAs encoding AGH precursors or putative AGH precursors have been cloned from three isopod (*2*, *4*) and one decapod species (*5*).

Similar to proinsulin, AGH is expressed as a 123-residue single precursor peptide chain consisting of B-chain (44 aa), C-peptide (50 aa), and A-chain (29 aa). Removing the C-peptide portion by processing at two Lys-Arg sites transforms it into a mature heterodimeric form which contains four disulfide bonds (Figure 1) (*2*, *6*). It has been shown that the mature AGH harbors an N-linked glycan at Asn^{A18} in the A-chain, and this glycan moiety was essential for conferring the biological activity (*6*). To study the relationship between biological activity and glycan structure, we sought to obtain an AGH that had a homogeneous glycan structure.

When large amounts of glycoprotein are required, recombinant protein expression systems are used. However, recombinant glycoproteins expressed in baculovirus, yeast, or mammalian cell systems usually have heterogeneous glycan structures, and it is difficult to obtain homogeneous materials from these systems. A chemical synthesis strategy provides a good solution for this problem.

In 1998, Muir et al. developed the expressed protein ligation method (*7*). In this method, a recombinant protein thioester prepared from a C-terminal intein tag-fusion protein was used as a building block and condensed with a peptide segment containing posttranslational modifications and/or unnatural amino acids that were prepared by the ordinary solid-phase peptide synthesis method. Up to now, several proteins such as phospho-, glyco-, and acyl-proteins have been synthesized using this method (*8*). Since the glycosylation site of AGH is in the C-terminal part of its precursor, this method may be useful for synthesizing AGH. In this study, in order to obtain AGH with a homogeneous glycan structure, we synthesized AGH using the expressed protein ligation method.

EXPERIMENTAL PROCEDURES

General. MALDI-TOF mass spectra were recorded with a Voyager-DE PRO spectrometer (Applied Biosystems, CA). Amino acid composition was determined with a LaChrom amino acid analyzer (Hitachi, Tokyo, Japan) after hydrolysis with 6 M HCl at 150 °C for 2 h in an evacuated sealed tube. Amino acid sequence analysis was performed with a protein sequencer 491 (Applied Biosystems). Circular dichroism (CD) spectra were measured with a Jasco J-820 spectropolarimeter (JASCO, Tokyo, Japan) at room temperature with a 1 mm path length cell. Protein concentration was determined from the absorbance at 280 nm using the extinction constants of Tyr and Trp residues as reported previously (*9*).

Expression and Purification of AGH(1−110) Thioester ***1***. Isolation and characterization of a cDNA encoding AGH of

[†]This work was supported in part by a Grant-in Aid for Scientific Research from the Ministry of Education, Sport, Science, and Technology of Japan (No. 20380069). We thank Tokai University for a Grant-in-Aid for High-Technology Research and the Japan Society for the Promotion of Science for a Grant-in-Aid for Creative Scientific Research (No. 17GS0420).

*To whom correspondence should be addressed: e-mail, katay@hotmail.co.jp (H.K.) and hojo@keyaki.cc.u-tokai.ac.jp (H.H.); phone/fax, +81-463-50-2075.

 Published on Web 01/21/2010

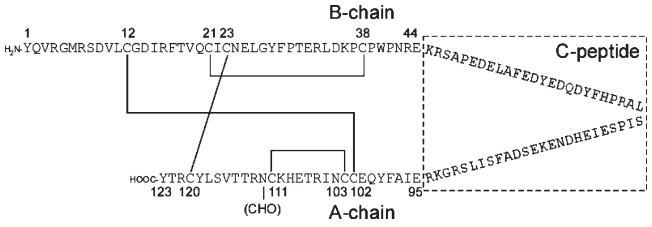


FIGURE 1: Primary structure of AGH. (CHO), N-glycosylation site.

A. vulgare have been described previously (*2*). A cDNA encoding AGH(1−110) was amplified by polymerase chain reaction (PCR) using the following set of primers: a forward primer (5′-TGGTC-ATATGTACCAGGTACGAGGTATGAGATCCGATGTG-3′) and a reverse primer (5′-GGTTGCTCTTCTGCATTTGTGC-TCTGTCCTAATATTGC-3′). The amplified PCR product was cloned into *Nde*I/*Sap*I sites of the pTWIN1 vector (New England Biolabs, MA) to obtain an expression plasmid, pTWIN1-AGH-(1−110).

Escherichia coli ER2566 competent cells were transformed with the expression plasmid and selected on LB plates containing ampicillin (50 μg/mL). Bacterial cells from a single colony were grown at 37 °C overnight in LB medium containing ampicillin and then diluted 50-fold with the same medium. The diluted medium was incubated at 37 °C for 90 min, and then isopropyl β-D-thiogalactoside (IPTG) was added to the culture at a final concentration of 1 mM. After an additional incubation for 3 h, bacterial cells were harvested by centrifugation. The bacterial cells were suspended in $^1/_{50}$th culture volume of phosphate-buffered saline (PBS; 150 mM NaCl, 10 mM phosphate buffer, pH 7.0), and the suspension was sonicated. The insoluble material was collected by centrifugation and washed twice with the same volume of PBS to obtain pure inclusion bodies.

After solubilizing the inclusion bodies with a $^1/_{100}$th culture volume of 8 M urea solution, the resultant solution was diluted using a 4-fold volume of dilution buffer to final conditions of 1.6 M urea/10% glycerol/50 mM sodium mercaptoethanesulfonate (MESNa)/100 mM Hepes (pH 7.5) and stirred at room temperature overnight to refold the intein tag part of the recombinant protein and to digest the tag by MESNa. Proteins in solution were precipitated by adding trichloroacetic acid (TCA) at the concentration of 10%, and the precipitate was washed twice with acetone. The desired product **1** was purified by gel filtration column chromatography on a Superdex 75 10/300 GL column (GE Healthcare UK Ltd., U.K.) with 50% acetonitrile/0.1% trifluoroacetic acid (TFA) as a solvent at a flow rate of 0.75 mL/min.

N^2-(9-Fluorenylmethoxycarbonyl)-N^4-(2-acetamido-3,4,6-tri-O-benzyl-2-deoxy-β-D-glucopyranosyl)-L-asparagine tert-Butyl Ester ***3***. Fmoc-Asp-OBu^t (1.6 g, 3.9 mmol) was dissolved in dichloromethane (6 mL), and *N*,*N*′-dicyclohexylcarbodiimide (DCC; 410 mg, 2.0 mmol) was added at 0 °C. After the mixture was stirred for 30 min at 0 °C, the precipitate was filtered off, and the solvent was removed in vacuo. The residue was dissolved in tetrahydrofuran (6 mL) and added to a mixture of 2-acetamido-3,4,6-tri-*O*-benzyl-2-deoxy-β-D-glucopyranosyl azide **2** (520 mg, 1.0 mmol) and Lindlar catalyst (300 mg) (*10*). The resulting solution was stirred overnight under a H_2 atmosphere (*11*). After the catalyst was removed on Celite, the solvent was removed in vacuo. The residue was purified by silica gel column chromatography with 95:5 $CHCl_3$−MeOH to give compound **3** (660 mg, 0.75 mmol, 75%): $[\alpha]_D$ +45.9° (*c* = 0.5 in $CHCl_3$); R_f 0.52 (95:5 $CHCl_3$:MeOH); ^{1}H NMR ($CDCl_3$−D_2O) δ 4.85 (d, 1H, *J* = 12.0 Hz, Ph-C$\underline{H}_2$), 4.80 (d, 1H, *J* = 9.5 Hz, H-1), 4.79 (d, 1H, *J* = 10.7 Hz, Ph-C$\underline{H}_2$), 4.62 (d, 2H, *J* = 12.0 Hz, Ph-C$\underline{H}_2$), 4.57 (d, 1H, *J* = 10.7 Hz, Ph-C$\underline{H}_2$), 4.48 (m, 1H, Asn α-H), 4.45 (d, 1H, *J* = 12.2 Hz, Ph-C$\underline{H}_2$), 4.40 (dd, 1H, *J* = 7.1, 10.3 Hz, Fmoc CH-C$\underline{H}_2$), 4.27 (t, 1H, *J* = 8.8 Hz, Fmoc CH-C$\underline{H}_2$), 4.20 (t, 1H, *J* = 7.2 Hz, Fmoc C$\underline{H}$-CH$_2$), 3.87 (t, 1H, *J* = 10.0 Hz, H-2), 3.78 (t, 1H, *J* = 9.3 Hz, H-4), 3.45 (dd, 1H, *J* = 9.0, 10.3 Hz, H-3), 2.79 (dd, 1H, *J* = 4.9, 16.4 Hz, Asn β-H), 2.66 (dd, 1H, *J* = 4.0, 16.6 Hz, Asn β-H), 1.71 (s, 3H, CH_3CO), 1.41 (s, 9H, *t*-Bu). Anal. Calcd for $C_{52}H_{57}N_3O_{10}$: C, 70.65; H, 6.50; N, 4.75; O, 18.10. Found: C, 70.67; H, 6.50; N, 4.70.

N^2-(9-Fluorenylmethoxycarbonyl)-N^4-(2-acetamido-3,4,6-tri-O-benzyl-2-deoxy-β-D-glucopyranosyl)-L-asparagine ***4***. Compound **3** (170 mg, 0.19 mmol) was dissolved in 90% TFA in dichrolomethane (4 mL), and the solution was stirred for 1 h. The solvent was removed in vacuo to obtain Fmoc-Asn-(GlcNAcBn$_3$)-OH **4** (120 mg, 0.14 mmol, 76%): $[\alpha]_D$ +10.1° (*c* = 0.5 in DMF); R_f 0.38 (90:10:1 $CHCl_3$:MeOH:AcOH); ^{1}H NMR (DMSO) δ 12.63 (br s, 1H, COO$\underline{H}$), 8.41 (d, 1H, *J* = 9.3 Hz, Asn γ-CON$\underline{H}$), 7.99 (d, 1H, *J* = 9.3 Hz, N$\underline{H}$-Ac), 7.48 (d, 1H, *J* = 8.3 Hz, Asn NH), 4.98 (t, 1H, *J* = 9.4 Hz, H-1), 4.70 (d, 1H, *J* = 11.0 Hz, Ph-C$\underline{H}_2$), 4.69 (d, 1H, *J* = 11.0 Hz, Ph-C$\underline{H}_2$), 4.65 (d, 1H, *J* = 11.0 Hz, Ph-C$\underline{H}_2$), 4.53 (d, 1H, *J* = 12.0 Hz, Ph-C$\underline{H}_2$), 4.52 (d, 1H, *J* = 11.0 Hz, Ph-C$\underline{H}_2$), 4.46 (d, 1H, *J* = 12.0 Hz, Ph-C$\underline{H}_2$), 4.40 (m, 1H, Asn α-H), 3.79 (m, 1H, H-2), 2.66 (dd, 1H, *J* = 5.4, 16.1 Hz, Asn β-H), 1.77 (s, 3H, CH_3CO). Anal. Calcd for $C_{48}H_{49}N_3O_{10}\cdot{}^1/_2H_2O$: C, 68.89; H, 6.02; N, 5.02. Found: C, 68.93; H, 5.96; N, 5.00.

Nonglycosylated AGH(111−123) ***5***. Starting from Fmoc-Tyr(Bu^t)-Wang resin (0.63 mmol/g, 397 mg), the peptide chain was elongated with an ABI 433A peptide synthesizer (Applied Biosystems) using FastMoc protocol, and Fmoc-Arg(Pbf)-Thr-(Bu^t)-Thr(Bu^t)-Val-Ser(Bu^t)-Leu-Tyr(Bu^t)-Cys(Trt)-Arg(Pbf)-Thr-(Bu^t)-Tyr(Bu^t)-Wang resin was obtained. Four-fifths of the resin was removed, and the remaining resin was used for additional peptide chain elongation with the peptide synthesizer. After the peptide chain elongation, H-Cys(Trt)-Asn(Trt)-Arg(Pbf)-Thr-(Bu^t)-Thr(Bu^t)-Val-Ser(Bu^t)-Leu-Tyr(Bu^t)-Cys(Trt)-Arg(Pbf)-Thr-(Bu^t)-Tyr(Bu^t)-Wang resin was obtained (246 mg). A part of the resin (100 mg) was treated with reagent K (TFA/phenol/water/thioanisole/ethanedithiol, 33:2:2:2:1, 1.5 mL) (*12*) for 2 h. TFA was removed with a nitrogen stream, and the peptide was precipitated by diethyl ether. The crude peptide was washed twice with ether, and AGH(111−123) **5** was purified by reversed-phase (RP) HPLC on a Mightysil RP-18 GP column (4.6 × 150 mm; Kanto, Japan) at a flow rate of 1 mL/min. Elution was performed with a 30 min linear gradient of 10−40% acetonitrile containing 0.1% TFA. The yield was 32.2 mg (20.4 μmol, 50%). MALDI-TOF mass: found, *m*/*z* 1579.4; calcd, 1579.7 for $(M + H)^+$. Amino acid analysis: $Asp_{1.00}Thr_{2.93}Ser_{0.95}Cys_{0.09}Val_{1.00}Leu_1Tyr_{2.07}Arg_{2.16}$.

Asn(GlcNAc)^{112}-AGH(111−123) ***6***. Fmoc-Arg(Pbf)-Thr-(Bu^t)-Thr(Bu^t)-Val-Ser(Bu^t)-Leu-Tyr(Bu^t)-Cys(Trt)-Arg(Pbf)-Thr-(Bu^t)-Tyr(Bu^t)-Wang resin (50 μmol equivalent) was treated with 20% piperidine/1-methyl-2-pyrrolidinone (NMP) and reacted with Fmoc-Asn(GlcNAcBn$_3$)-OBt, which was prepared from Fmoc-Asn(GlcNAcBn$_3$)-OH **4**, 1 M DCC/NMP, and 1 M 1-hydroxybenzotriazole (HOBt)/NMP. After the reaction at 50 °C for 1 h, the resin was washed with NMP, and the N-terminal Fmoc group was cleaved with 20% piperidine/NMP treatment. Fmoc-Cys(Trt)-OH was then condensed in the same manner as

Fmoc-Asn(GlcNAcBn$_3$)-OH. After cleavage of the Fmoc group by 20% piperidine/NMP treatment, H-Cys(Trt)-Asn-(GlcNAcBn$_3$)-Arg(Pbf)-Thr(Bu^t)-Thr(Bu^t)-Val-Ser(Bu^t)-Leu-Tyr-(Bu^t)-Cys(Trt)-Arg(Pbf)-Thr(Bu^t)-Tyr(Bu^t)-Wang resin was obtained (207 mg). A part of the resin (100 mg) was treated with reagent K (1.5 mL) for 2 h. TFA was removed with a nitrogen stream, and the peptide was precipitated by ether. The crude peptide was washed twice with ether and then dissolved in a mixture of TFA/dimethyl sulfide/*m*-cresol (5:3:1, 0.9 mL). Triflic acid (0.1 mL) was added to the solution, and the reaction mixture was left at −15 °C for 2 h (*13*−*16*). The crude peptide was precipitated and washed twice with ether and separated by RP-HPLC under the same conditions as for the nonglycosylated AGH(111−123) to give Asn(GlcNAc)^{112}-AGH(111−123) **6** (43.1 mg, 24.2 μmol, 37%). MALDI-TOF mass: found, *m*/*z* 1782.9; calcd, 1782.8 for $(M + H)^+$. Amino acid analysis: $Asp_{1.03}$-$Thr_{2.92}Ser_{0.96}Cys_{0.23}Val_{1.28}Leu_1Tyr_{2.06}Arg_{2.14}$.

Asn[(*GalGlcNAcMan*)$_2$*ManGlcNAc*$_2$]^{112}*-AGH*(*111*−*123*) ***7***. Fmoc-Asn[(GalGlcNAcMan)$_2$ManGlcNAc$_2$]-OH was a kind gift from Otsuka Chemical Co. Ltd., Japan. Fmoc-Arg(Pbf)-Thr(Bu^t)-Thr(Bu^t)-Val-Ser(Bu^t)-Leu-Tyr(Bu^t)-Cys(Trt)-Arg(Pbf)-Thr(Bu^t)-Tyr(Bu^t)-Wang resin (40 μmol equivalent) was treated with 20% piperidine/NMP. Fmoc-Asn[(GalGlcNAcMan)$_2$Man-GlcNAc$_2$]-OH (50 μmol) was activated with 3-(diethoxyphosphoryloxy)-3*H*-benzo[*d*][1,2,3]-triazin-4-one (DEPBT) in NMP and reacted with the resin (*17*, *18*). After reaction at 50 °C for 2 h, the resin was washed with NMP, and the N-terminal Fmoc group was cleaved by 20% piperidine/NMP treatment. Fmoc-Cys(Trt)-OH was then condensed by DCC/HOBt activation to obtain H-Cys(Trt)-Asn[(GalGlcNAcMan)$_2$ManGlcNAc$_2$]-Arg(Pbf)-Thr(Bu^t)-Thr(Bu^t)-Val-Ser(Bu^t)-Leu-Tyr(Bu^t)-Cys(Trt)-Arg(Pbf)-Thr(Bu^t)-Tyr(Bu^t)-Wang resin (174 mg). A part of the resin (44 mg) was treated with reagent K (1.0 mL) for 2 h. TFA was removed with a nitrogen stream, and the peptide was precipitated with ether. The crude peptide was washed twice with ether, and the product was purified by RP-HPLC under the same conditions as for the nonglycosylated AGH(111−123) to give Asn[(GalGlcNAcMan)$_2$ManGlcNAc$_2$]^{112}-AGH(111−123) **7** (0.90 mg, 0.28 μmol, 2.8%). MALDI-TOF mass: found, *m*/*z* 3202.3; calcd, 3202.3 for $(M + H)^+$. Amino acid analysis: $Asp_{0.84}$ $Thr_{2.64}Ser_{0.92}Cys_{0.36}Val_{2.00}Leu_1Tyr_{1.87}Arg_{1.93}$.

Nonglycosylated AGH ***10***. The N-terminal peptide thioester **1** (1 mg) and nonglycosylated C-terminal peptide **5** (1 mg) were mixed in a 1.5 mL microtube and dissolved in 100 μL of 100 mM phosphate buffer (pH 7.0) containing 6 M guanidine hydrochloride. The coupling reaction was initiated by adding 4-mercaptophenylacetic acid (MPAA) at the concentration of 2%. After reaction at room temperature for 3 days, the reaction mixture was desalted by gel filtration column chromatography on a Superdex 75 10/300 GL column with 50% acetonitrile/0.1% TFA as a solvent at a flow rate of 0.75 mL/min. After lyophilization, the resultant material was dissolved in 1 mL of 8 M urea/90 mM glutathione (reduced form, GSH). The solution was then diluted to the final conditions of 1.3 M urea/10% glycerol/5 mM GSH/200 mM Tris (pH 8.5), 18 mL, and cooled at 4 °C. The oxidized form of glutathione (GSSG) was then added to the solution at a concentration of 1 mM, and the solution was stirred for 3 days at 4 °C. The folded AGH **9** was purified by RP-HPLC on a PROTEIN-RP column (YCM, Japan). Elution was performed with a 30 min linear gradient of 20−50% acetonitrile containing 0.1% TFA. After lyophilization, the resultant peptide was dissolved in 100 μL of 100 mM Tris (pH 6.8), and then 0.1 unit of metalloendopeptidase (Seikagaku Kogyo, Japan) was added. After 2 h incubation at 37 °C, product **10** was purified by RP-HPLC in the same manner as described above. The yield was 14 μg (2.0%). MALDI-TOF mass: found, *m*/*z* 9161.8; calcd, 9162.2 (average) for $(M + H)^+$.

Asn(*GlcNAc*)^{112}*-AGH* ***15***. The N-terminal peptide thioester **1** (1 mg) and monosaccharyl C-terminal peptide **6** (1 mg) were used for the coupling reaction. The conditions for coupling, folding, and processing reactions were the same as those for **10**. The isolated yield of **15** was 57 μg (8.0%). MALDI-TOF mass: found, *m*/*z* 9366.4; calcd, 9363.4 (average) for $(M + H)^+$.

Asn[(*GalGlcNAcMan*)$_2$*ManGlcNAc*$_2$]^{112}*-AGH* ***16***. The N-terminal peptide thioester **1** (1 mg) and nonasaccharyl C-terminal peptide **7** (0.9 mg) were used for the coupling reaction. The conditions for coupling, folding, and processing reactions were the same as those for **10**. The isolated yield of **16** was 14 μg (1.6%). MALDI-TOF mass: found, *m*/*z* 10784.6; calcd, 10784.7 (average) for $(M + H)^+$.

Bioassay. AGH activity was determined using an *in vivo* bioassay as described (*19*, *20*). In brief, samples (10 ng each of the synthetic AGH or 0.5 AG equivalent of AG extract) were added to a small amount of bovine serum albumin, dried under vacuum, and dissolved in saline solution. They were injected into 10 young *A. vulgare* females. After molting, the animals were examined under a stereoscopic microscope for elongation of the endopodites of the first pair of abdominal legs, a criterion for masculinization.

Trypsin Digestion. Peptides **10** and **15** (10 μg each) were dissolved in 100 mL of 50 mM Hepes buffer (pH 7.0), and 1 μL of TPCK-treated trypsin (Sigma, MO) aqueous solution (1 mg/mL) was added. The solution was incubated at 37 °C for 24 h, and the tryptic digests were separated by RP-HPLC on a Mightysil RP-18 GP column. Elution was performed with a 30 min linear gradient of 10−40% acetonitrile containing 0.1% TFA.

Endoproteinase Glu-C Digestion. The tryptic digest was dissolved in 100 μL of 100 mM Tris buffer (pH 6.8), and 0.5 μL of endoproteinase Glu-C (Sigma) aqueous solution (1 mg/mL) was added. The solution was incubated at 37 °C for 24 h, and the digests were separated by RP-HPLC in the same manner as that for the tryptic digests.

[*Asn*(*GlcNAcBn*$_3$)^{112},*Cys*(*MeOBn*)^{102},*Cys*(*Acm*)^{120}]*-AGH A-Chain* ***17***. Starting from Fmoc-Tyr(Bu^t)-OCH$_2$-C$_6$H$_4$-OCH$_2$-C$_6$H$_4$-resin (520 mg, 250 μmol), the peptide chain was elongated with an ABI 433A peptide synthesizer using FastMoc protocol, and Arg(Pbf)-Thr(Bu^t)-Thr(Bu^t)-Val-Ser(Bu^t)-Leu-Tyr(Bu^t)-Cys(Acm)-Arg(Pbf)-Thr(Bu^t)-Tyr(Bu^t)-OCH$_2$-C$_6$H$_4$-OCH$_2$-C$_6$H$_4$-resin was obtained. A part of the resin (70 μmol) was removed and reacted with Fmoc-Asn(GlcNAcBn$_3$)-OBt, which was prepared from Fmoc-Asn(GlcNAcBn$_3$)-OH (83 mg, 100 μmol), 1 M DCC/NMP (110 μmol), and 1 M HOBt/NMP (110 μmol). After 1.5 h, the resin was washed with NMP and acetylated with 10% Ac$_2$O and 5% DIEA in NMP for 10 min. Chain elongation was continued using the peptide synthesizer, and Glu(OBu^t)-Ile-Ala-Phe-Tyr(Bu^t)-Gln(Trt)-Glu(OBu^t)-Cys(MeOBn)-Cys(Trt)-Asn(Trt)-Ile-Arg(Pbf)-Thr(Bu^t)-Glu(OBu^t)-His(Trt)-Lys(Boc)-Cys(Trt)-Asn(GlcNAcBn$_3$)-Arg(Pbf)-Thr(Bu^t)-Thr(Bu^t)-Val-Ser-(Bu^t)-Leu-Tyr(Bu^t)-Cys(Acm)-Arg(Pbf)-Thr(Bu^t)-Tyr(Bu^t)-OCH$_2$-C$_6$H$_4$-OCH$_2$-C$_6$H$_4$-resin (444 mg) was obtained. A part of the resin (108 mg) was treated with reagent K (1 mL) for 2 h. TFA was removed with a nitrogen stream, and the peptide was precipitated with ether. The peptide was washed twice with ether and extracted with 50% aqueous acetonitrile (10 mL). The

solution was added dropwise into 0.1 M ammonium acetate containing 6 M guanidine hydrochloride (200 mL), and the solution was gently stirred for 3 days. Acetic acid was added, and the product was purified by RP-HPLC to obtain peptide **17** (2.7 μmol). The dibenzylated peptides were also collected (0.61 μmol). MALDI-TOF mass: found, m/z 4207.4; calcd, 4206.9 for $(M + H)^+$. Amino acid analysis: $Asp_{2.09}Thr_{3.79}Ser_{0.96}Glu_{4.25}Ala_1{}^1/_2Cys_{2.46}Val_{1.10}Ile_{1.74}Leu_{1.06}Tyr_{2.86}Phe_{0.92}Lys_{0.99}His_{0.97}Arg_{3.02}$.

[*Asn*(*GlcNAc*)^{112},*Cys*(*SPy*)^{102},*Cys*(*Acm*)^{120}]-*AGH A-Chain* ***18***. Peptide **17** (730 nmol), DPDS (6.4 mg, 29 μmol), and thioanisole (60 μL) were dissolved in TFA (600 μL), and TfOH (30 μL) was added at −10 °C. After the solution was kept for 5 min at −10 °C, the product was precipitated with ether and purified by RP-HPLC to give peptide **18** (299 nmol, 41%). MALDI-TOF mass: found, m/z 3925.4; calcd, 3925.7 for $(M + H)^+$. A major peak corresponding to the depyridine sulfenylated peptide, which may be derived from decomposition during the mass measurement, was also observed: found, m/z 3816.5; calcd, 3816.7 for [Asn(GlcNAc)^{112},Cys(Acm)^{120}]-AGH A-chain $(M + H)^+$. Amino acid analysis: $Asp_{2.17}Thr_{3.83}Ser_{0.99}Glu_{4.64}Ala_1{}^1/_2Cys_{1.55}Val_{0.98}Ile_{1.83}Leu_{1.00}Tyr_{2.88}Phe_{0.94}Lys_{1.05}His_{1.00}Arg_{3.15}$.

[*Cys*(*MeOBn*)^{12},*Cys*(*Acm*)^{23}]-*AGH B-Chain* ***19***. Starting from Fmoc-Glu(OBu^t)-Wang resin (0.50 g, 275 μmol), the peptide chain was elongated using the peptide synthesizer, and Tyr(Bu^t)-Gln(Trt)-Val-Arg(Pbf)-Gly-Met-Arg(Pbf)-Ser(Bu^t)-Asp(OBu^t)-Val-Leu-Cys(MeOBn)-Gly-Asp(OBu^t)-Ile-Arg(Pbf)-Phe-Thr(Bu^t)-Val-Gln(Trt)-Cys(Trt)-Ile-Cys(Acm)-Asn(Trt)-Glu(OBu^t)-Leu-Gly-Tyr(Bu^t)-Phe-Pro-Thr(Bu^t)-Glu(OBu^t)-Arg(Pbf)-Leu-Asp(OBu^t)-Lys(Boc)-Pro-Cys(Trt)-Pro-Trp(Boc)-Pro-Asn(Trt)-Arg(Pbf)-Glu(OBu^t)-Wang resin (2.2 g) was obtained. An aliquot of the resin (50 mg) was treated with reagent K (500 μL) for 2 h at room temperature. The peptide precipitated with ether was dissolved in 50% aqueous acetonitrile containing 0.1% TFA and diluted with 0.1 M ammonium acetate (50 mL) containing 6 M guanidine hydrochloride. The pH was adjusted to 8.0 with aqueous ammonia, and the solution was stirred for 3 days. The main peak was isolated by RP-HPLC to give peptide **19** (368 nmol, 5.9% yield based on the content of Glu residue in the starting resin). MALDI-TOF mass: found, m/z 5397.7; calcd, 5397.2 (average) for $(M + H)^+$. Amino acid analysis: $Asp_{5.10}Thr_{1.96}Ser_{0.91}Glu_{5.10}Pro_{3.94}Gly_{3.00}{}^1/_2Cys_{2.50}Val_{2.79}Met_{0.72}Ile_{1.69}Leu_{3.14}Tyr_{1.81}Phe_{2.13}Lys_{1.01}Arg_{4.74}$.

[*Cys*(*Acm*)^{23}]-*AGH B-Chain* ***20***. Peptide **19** (790 nmol), *m*-cresol (15 μL), and thioanisole (30 μL) were dissolved in TFA (210 μL), and TfOH (26 μL) was added at 0 °C. After the solution was kept at 0 °C for 20 min with occasional mixing, the product was precipitated with ether and purified by RP-HPLC to give peptide **20** (380 μmol, 48%). MALDI-TOF mass: found, m/z 5273.4; calcd, 5273.5 for $(M + H)^+$. Amino acid analysis: $Asp_{5.03}Thr_{1.73}Ser_{0.87}Glu_{5.27}Pro_{4.11}Gly_{2.71}{}^1/_2Cys_{2.46}Val_{2.43}Met_{0.84}Ile_{1.24}Leu_{2.80}Tyr_{1.83}Phe_2Lys_{1.06}Arg_{4.61}$.

[*Asn*(*GlcNAc*)^{112},*Cys*(*Acm*)$^{23,\ 120}$]-*AGH* ***21***. Peptide **20** (134 nmol) in 50% acetonitrile (0.8 mL) was added dropwise to a mixture of peptide **18** (224 nmol), 0.1 M NH_4HCO_3 (0.8 mL), and acetonitrile (0.4 mL) within 1 min. The reaction was quenched by adding acetic acid, and the product was purified by RP-HPLC to give peptide **21** (80 nmol, 60%). MALDI-TOF mass: found, m/z 9093.3; calcd, 9093.4 (average) for $(M + H)^+$. Amino acid analysis: $Asp_{6.82}Thr_{5.22}Ser_{1.67}Glu_{8.65}Pro_{4.41}Gly_{3.06}Ala_1{}^1/_2Cys_{2.26}Val_{3.99}Met_{1.08}Ile_{3.51}Leu_{3.96}Tyr_{5.22}Phe_{3.09}Lys_{2.04}His_{1.04}Arg_{7.12}$.

[*Asn*(*GlcNAc*)^{112}]-*AGH* ***22***. Peptide **21** (80 nmol) was dissolved in 20% aqueous MeOH (1.0 mL) and added to MeOH (5.0 mL) containing 20 mM I_2 in MeOH (200 μL) and 6 M HCl (67 μL) within 7 min with mixing. After the resulting solution was stirred for another 15 min, the reaction was terminated by adding ascorbic acid. The product was purified by cation-exchange column chromatography (TSKgel CM-3SW, 7.5 mm × 750 mm) using a linear gradient from 20 mM sodium phosphate (pH 6.0) containing 40% MeCN to the same buffer containing 0.5 M NaCl in 20 min at a flow rate of 0.5 mL/min and then desalted by RP-HPLC to give peptide **22** (31 nmol, 39%). MALDI-TOF mass: found, m/z 8949.8; calcd, 8949.2 (average) for $(M + H)^+$. Amino acid analysis: $Asp_{7.00}Thr_{5.42}Ser_{1.79}Glu_{8.92}Pro_{4.46}Gly_{3.16}Ala_1{}^1/_2Cys_{2.66}Val_{4.05}Met_{0.92}Ile_{3.75}Leu_{4.16}Tyr_{5.43}Phe_{3.38}Lys_{2.08}His_{1.01}Arg_{7.20}$.

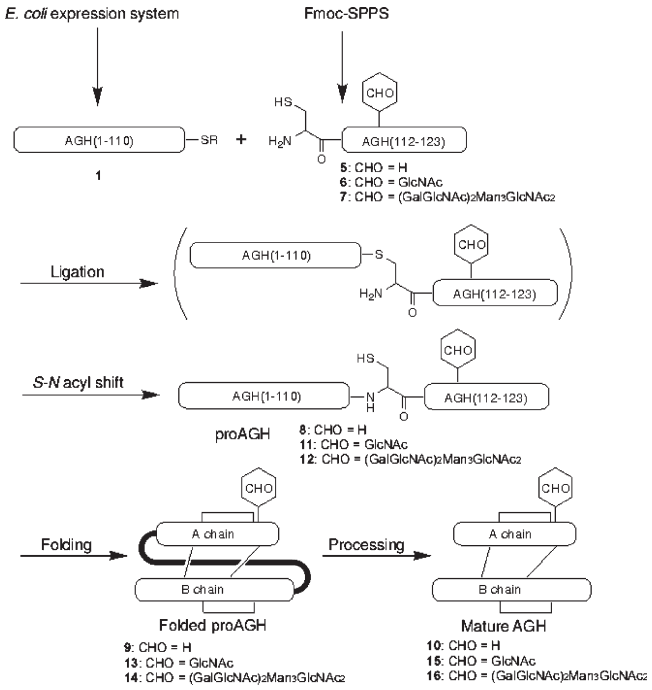
[*Cys*(*MeOBn*)^{102},*Cys*(*Acm*)^{120}]-*AGH A-Chain* ***24***. Starting from Fmoc-Tyr(Bu^t)-OCH$_2$-C$_6$H$_4$-OCH$_2$-C$_6$H$_4$-resin (510 mg, 245 μmol), the peptide chain was elongated with the peptide synthesizer using FastMoc protocol, and Glu(OBu^t)-Ile-Ala-Phe-Tyr(Bu^t)-Gln(Trt)-Glu(OBu^t)-Cys(MeOBn)-Cys(Trt)-Asn(Trt)-Ile-Arg(Pbf)-Thr(Bu^t)-Glu(OBu^t)-His(Trt)-Lys(Boc)-Cys(Trt)-Asn(Trt)-Arg(Pbf)-Thr(Bu^t)-Thr(Bu^t)-Val-Ser(Bu^t)-Leu-Tyr(Bu^t)-Cys(Acm)-Arg(Pbf)-Thr(Bu^t)-Tyr(Bu^t)-OCH$_2$-C$_6$H$_4$-OCH$_2$-C$_6$H$_4$-resin (2.1 g) was obtained. An aliquot of the resin (100 mg) was treated with 1 mL of reagent K for 2 h at room temperature. The peptide precipitated by ether was dissolved in 50% aqueous acetonitrile containing 0.1% TFA (10 mL) and diluted with 0.1 M ammonium acetate (90 mL) containing 6 M guanidine hydrochloride. The pH was adjusted to 8.0 with aqueous ammonia, and the solution was stirred for 3 days. The main peak was isolated by RP-HPLC to give peptide **24** (1.1 μmol, 4.1 mg, 9.3% yield). MALDI-TOF mass: found, m/z 3733.2; calcd, 3733.7 for $(M + H)^+$. Amino acid analysis: $Asp_{1.96}Thr_{3.48}Ser_{0.90}Glu_{4.05}Ala_{1.00}{}^1/_2Cys_{3.18}Val_{0.90}Ile_{1.71}Leu_{0.99}Tyr_{2.96}Phe_{0.95}Lys_{0.94}His_{0.94}Arg_{2.75}$.

[*Cys*(*SPy*)^{102},*Cys*(*Acm*)^{120}]-*AGH A-Chain* ***25***. Peptide **24** (15 mg, 2.0 μmol), DPDS (16 mg, 72.6 μmol), and thioanisole (60 μL) were dissolved in TFA (540 μL), and the solution was kept at room temperature for 2.5 h. The product was precipitated with ether and purified by RP-HPLC to give peptide **25** (5.7 mg, 0.93 μmol, 48%). MALDI-TOF mass: found, m/z 3722.9; calcd, 3722.6 for $(M + H)^+$. Amino acid analysis: $Asp_{2.20}Thr_{3.90}Ser_{0.99}Glu_{4.73}Ala_{1.00}{}^1/_2Cys_{1.57}Val_{0.97}Ile_{1.86}Leu_{1.06}Tyr_{3.17}Phe_{1.07}Lys_{1.13}His_{1.07}Arg_{3.17}$.

Cys(*Acm*)$^{23,\ 120}$-*AGH* ***26***. Peptide **20** (303 nmol) in 50% acetonitrile (1.0 mL) was added dropwise to the solution of **25** (312 nmol) in 0.1 M NH_4HCO_3 (0.8 mL) and acetonitrile (0.4 mL) within 1 min. The reaction was quenched by adding acetic acid, and the product was purified by RP-HPLC to give peptide **26** (286 nmol, 94%). MALDI-TOF mass: found, m/z 8890.5; calcd, 8890.2 (average) for $(M + H)^+$. Amino acid analysis: $Asp_{6.95}Thr_{5.24}Ser_{1.72}Glu_{9.42}Pro_{4.02}Gly_{2.66}Ala_1{}^1/_2Cys_{3.67}Val_{3.35}Met_{0.89}Ile_{2.93}Leu_{3.71}Tyr_{4.72}Phe_{2.96}Lys_{2.08}His_{0.96}Arg_{7.49}$.

Nonglycosylated AGH ***23***. Peptide **26** (60 nmol) was dissolved in distilled water (0.5 mL) and added dropwise to MeOH (2.0 mL) containing 20 mM I_2 in MeOH (60 μL) and 6 M HCl (20 μL) within 5 min with mixing. After the resulting solution was stirred for another 15 min, the reaction was terminated by adding aqueous ascorbic acid solution. The product was purified by cation-exchange column chromatography (TSKgel CM-3SW, 7.5 mm × 75 mm) using a linear gradient from 20 mM sodium phosphate (pH 6.0) containing 40%

Scheme 1: Synthesis Strategy for AGH by the Expressed Protein Ligation Method



acetonitrile to the same buffer containing 0.5 M NaCl in 20 min at a flow rate of 0.5 mL/min and desalted by RP-HPLC to give peptide **23** (20 nmol, 33%). MALDI-TOF mass: found, *m*/*z* 8748.6, calcd, 8746.0 (average) for $(M + H)^+$. Amino acid analysis: $Asp_{6.93}Thr_{5.24}Ser_{1.70}Glu_{9.55}Pro_{4.53}Gly_{2.64}Ala_1{}^1/_2Cys_{4.67}Val_{3.51}Met_{0.96}Ile_{3.10}Leu_{3.77}Tyr_{4.85}Phe_{2.99}Lys_{2.12}His_{0.98}Arg_{7.66}$.

RESULTS

*Preparation of the N-Terminal Peptide Thioester **1***. In order to synthesize AGH using the expressed protein ligation method (Scheme 1), the N-terminal peptide thioester was prepared using a bacterial expression system. A recombinant AGH-(1−110) with a C-terminal intein−chitin binding protein tag was expressed in an *E. coli* expression system. Furthermore, an additional Met residue was attached at the N-terminus of the fusion protein. Unfortunately, most of the recombinant protein was expressed as inclusion bodies, and only a small amount was detected in the soluble fraction after disrupting cells by sonication. No drastic change in expression level in the soluble fraction was observed when the expression was induced under milder conditions, such as low temperature and low isopropyl *β*-D-thiogalactoside (IPTG) concentration. Therefore, we decided to obtain the desired product from the inclusion bodies by an *in vitro* refolding method.

The inclusion bodies were dissolved and denatured in an 8 M urea solution, and the intein-tag portion was refolded by 5-fold dilution under neutral pH. Simultaneously, the tag was cleaved off by adding a thiol compound, sodium mercaptoethanesulfonate, to obtain an AGH(1−110) thioester. The reaction mixture was separated using gel filtration column chromatography as shown in Figure 2. The desired product **1** (∼14 kDa) was eluted just after the peak derived from the tag portion (∼24 kDa), and the yield was 6.8 mg from 400 mL culture.

Preparation of the C-Terminal Segment. The Asn unit bearing a GlcNAc moiety was prepared as shown in Scheme 2. 2-Acetamido-3,4,6-tri-*O*-benzyl-2-deoxy-*β*-D-glucopyranosyl azide

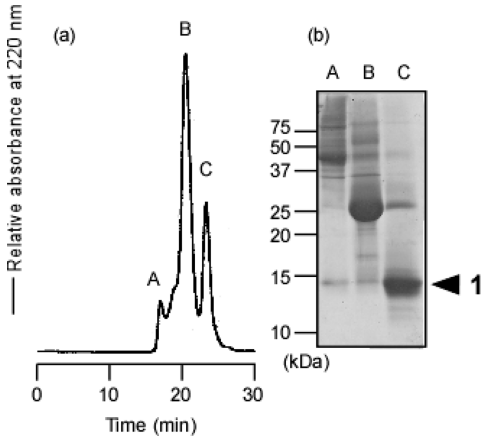
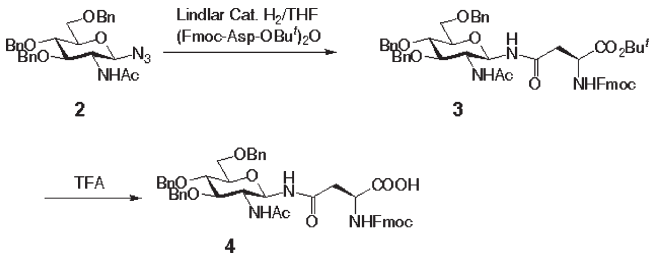


FIGURE 2: Preparation of the N-terminal peptide thioester **1**. (a) Gel filtration chromatogram (GFC) after the cleavage reaction of the recombinant intein fusion protein. (b) SDS−PAGE analysis of the peaks of GFC. Lane symbols indicate the peaks of GFC in panel a.

Scheme 2: Synthesis Route for Fmoc-Asn(GlcNAcBn$_3$)-OH



2 (*10*) was reduced and condensed using an anhydride of Fmoc-Asp-OBu^t in situ by Matsuo's method (*11*). The Bu^t ester was then cleaved with trifluoroacetic acid (TFA) to obtain Fmoc-Asn(GlcNAcBn$_3$)-OH **4**.

Nonglycosylated C-terminal segment **5** was synthesized by ordinary Fmoc chemistry. To synthesize monosaccharyl peptide **6**, Asn derivative **4** was introduced instead of Fmoc-Asn(Trt)-OH in the synthesis of **5** by the *N*,*N*′-dicyclohexylcarbodiimide (DCC)/1-hydroxybenzotriazole (HOBt) method. The protected peptide resins were treated with reagent K (*12*) for 2 h at room temperature, and benzyl groups on the sugar moiety were cleaved by low-acidity TfOH treatment (*13*−*16*). The nonasaccharyl peptide **7** was also synthesized and purified in the same manner as that for **5**, except for Asn^{112} in which Fmoc-Asn[(GalGlcNAcMan)$_2$ManGlcNAc$_2$] was introduced by 3-(diethoxyphosphoryloxy)-3*H*-benzo[*d*][1,2,3]-triazin-4-one (DEPBT) as a condensation reagent (*17*, *18*).

Native Chemical Ligation, Folding, and Processing. To synthesize nonglycosylated linear AGH(1−123) **8**, the N- and C-terminal segments, **1** and **5**, were condensed by the native chemical ligation reaction (*21*). The reaction was monitored by SDS−PAGE and was almost complete within 3 days (Figure 3). Even when the reaction continued for 7 days, the N-terminal segment was not completely consumed, suggesting that the thioester moiety was partially hydrolyzed. Since the product could not be separated from the hydrolyzed N-terminal segment, AGH(1−110), by reversed-phase (RP)-HPLC, we separated it from other additives by gel-filtration column chromatography and used it for the next reaction without further purification. Disulfide bonds were formed by oxidation in a redox buffer containing reduced and oxidized forms of glutathione, and folded propeptide **9** was purified by RP-HPLC (Figure 4).

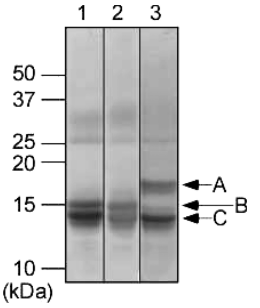


FIGURE 3: SDS−PAGE analysis of the ligation products: lane 1, nonglycosylated AGH; lane 2, monosaccharyl AGH; lane 3, nonasaccharyl AGH. A, the position of **12**; B, the positions of **8** and **11**; C, the positions of **1** and hydrolyzed material of **1**.

The C-peptide portion of **9** was removed by metalloendopeptidase treatment at 37 °C for 2 h, and mature AGH **10** was obtained (Figure 5). The isolated yield based on the amount of the N-terminal segment used for the peptide coupling reaction was 2.0%.

Monosaccharyl (**11**) and nonasaccharyl (**12**) propeptides were also synthesized by the native chemical ligation and were folded in the same manner as for the nonglycosylated peptide **9**, which provided folded propeptides **13** and **14**. The C-peptide was cleaved off in the same manner as for **10**, resulting in heterodimeric monosaccharyl (**15**) and nonasaccharyl (**16**) AGHs. Their isolated yields were 8.0% and 1.6%, respectively.

Circular dichroism (CD) spectra of the synthetic AGHs were similar regardless of their glycan structures (Figure 6). In addition, the spectral patterns showed that these AGHs possessed α-helical structures, which were similar to those of other insulin-family peptides (*22*, *23*).

Biological Activity and Confirmation of Folding. Biological activities of semisynthetic AGHs were assessed by *in vivo* bioassay for AGH. Unexpectedly, no sex reversal was observed in the female animals after an injection (200 ng/individual) of **10** or **15**, although injection of an AG extract (0.6 AG equivalent) induced masculinization (nine of nine individuals were masculinized). Since it has been reported that the native AGH isolated from the androgenic gland showed the activity at an injected dose of 0.1 ng/individual (*6*), it was likely that the synthetic AGHs did not have proper conformations.

In order to confirm folding, **10** and **15** were digested with endoproteases, and disulfide linkages of the digests were determined by mass spectral and N-terminal amino acid sequence analyses. First, AGHs were digested with TPCK-treated trypsin, and the digests were separated by RP-HPLC. On each HPLC chromatogram, two major peaks were observed. The analytical data for the fragments are summarized in Figure 7. These results indicated that each fragment consisted of three peptide chains connected by two disulfide bonds. When trypsin-2 fragment, a tryptic digest of nonglycosylated AGH **10**, was directly analyzed with a protein sequencer, di-PTH-Cys_2 was detected at cycles 5 and 7, establishing the arrangements of disulfide bonds between Cys^{21} and Cys^{38} and between Cys^{23} and Cys^{120}. The sequence analysis for trypsin-2′, one of the tryptic digests of monosaccharyl AGH **15**, gave the same result as trypsin-2. To determine the other two disulfide linkages, trypsin-1 and -1′ fragments were further digested with endoproteinase Glu-C. After digestion, fragments containing disulfide bonds were purified by RP-HPLC and analyzed with a protein sequencer. In the analysis of GluC-1, a digest of trypsin-1, di-PTH-Cys_2 was detected at cycles 1 and 5,

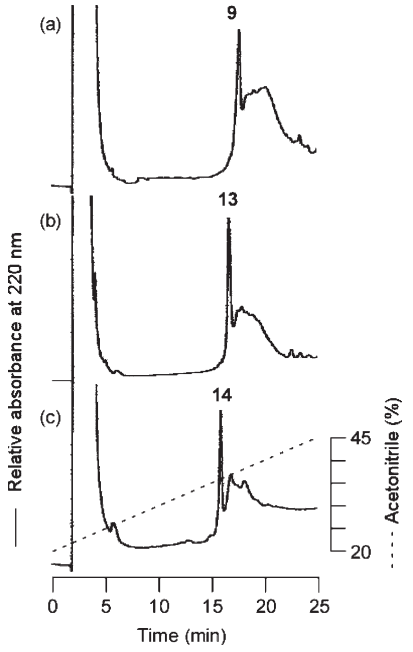


FIGURE 4: RP-HPLC elution profiles after the refolding reaction: (a) nonglycosylated AGH **9**; (b) monosaccharyl AGH **13**; (c) nonasaccharyl AGH **14**.

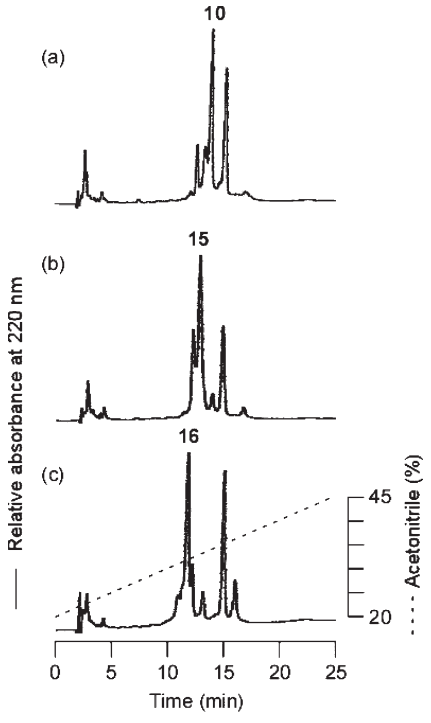


FIGURE 5: RP-HPLC elution profiles after metalloendopeptidase treatment: (a) nonglycosylated AGH **10**; (b) monosaccharyl AGH **15**; (c) nonasaccharyl AGH **16**.

but not at cycle 2, establishing the arrangements of disulfide bonds between Cys^{12} and Cys^{103} and between Cys^{102} and Cys^{111}. GluC-1′, a digest of trypsin-1′, gave the same results as GluC-1. These disulfide bond arrangements were not identical to that of recombinant AGH prepared in a baculovirus expression system, which had disulfide bonds between Cys^{12}-Cys^{102}, Cys^{21}-Cys^{38}, Cys^{23}-Cys^{120}, and Cys^{103}-Cys^{111} (*6*).

To obtain an AGH having correct disulfide bonds, we carried out refolding reactions under different conditions (reduced/

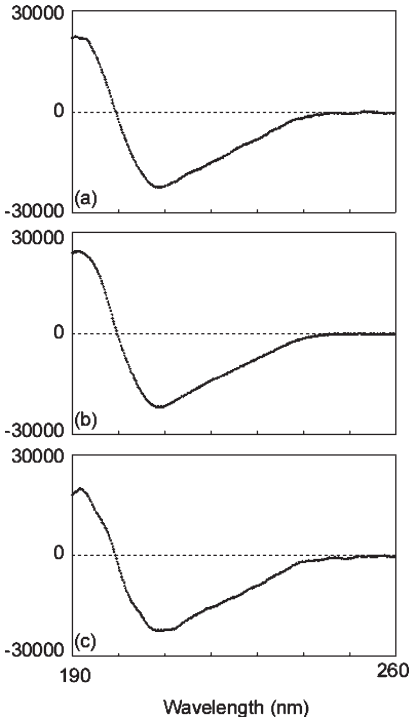


FIGURE 6: Circular dichroism spectra: (a) semisynthetic nonglycosylated AGH **10**; (b) semisynthetic monosaccharyl AGH **15**; (c) semisynthetic nonasaccharyl AGH **16**.

oxidized glutathione concentrations, pH, and additives) and also done by the stepwise dialysis method. However, these reactions gave the same results, and we could not obtain an AGH having native conformation.

Total Chemical Synthesis of AGH. To confirm that biological inactivity of the semisynthetic AGH was due to the disulfide isomerization, total chemical synthesis with selective disulfide bond formation was performed. The synthetic procedure is shown in Scheme 3. The A-chain with a GlcNAc moiety was synthesized using a peptide synthesizer. First, to form the intrachain disulfide bond, Cys^{103} and Cys^{111} were protected by a trityl (Trt) group. Cys^{102} was protected by a methoxybenzyl (MeOBn) group for the conversion to a pyridylsulfenyl (SPy) group by dipyridyl disulfide (DPDS). To form the final disulfide bond by iodine oxidation, Cys^{120} was protected by an acetamidomethyl (Acm) group. The introduction of Asn^{112} was accomplished using Fmoc-Asn($GlcNAcBn_3$) **4** by the DCC/HOBt method. The protected peptide resin was treated with reagent K for 2 h at room temperature. This treatment removed a part of the benzyl groups on the GlcNAc moiety.

Before purification, the crude peptide was air-oxidized in acetate buffer (pH 8) containing 6 M guanidine hydrochloride for 3 days to form an intrachain disulfide bond between Cys^{103} and Cys^{111}. The solution was loaded on an HPLC column, and A-chain congeners with tri- and dibenzyl GlcNAc were collected separately. The combined yield of these peptides was 19% based on the amount of Tyr on the initial resin. Peptides **17** were further treated with TfOH-TFA in the presence of DPDS for 5 min at −10 °C (*24*), in order to deprotect the MeOBn group of Cys^{102} and benzyl groups of carbohydrate moiety and also to introduce the SPy group into Cys^{102}. After purification, peptide **18** was obtained in 41% yield.

The B-chain was also prepared with a peptide synthesizer. After cleavage from the protected resin with reagent K, the crude peptide was air-oxidized and purified by RP-HPLC to obtain peptide **19** in 5.9% yield based on the amount of Glu on the initial resin. The peptide was then treated with TfOH−TFA at 0 °C for 20 min to remove the MeOBn group of Cys^{12} and was purified by RP-HPLC, obtaining peptide **20** in 48% yield.

The first interchain disulfide bond between Cys^{12} and Cys^{102} was constructed as follows. Peptide **20** dissolved in 50% aqueous acetonitrile was added dropwise into a solution of peptide **18** in 67 mM NH_4HCO_3 and 33% acetonitrile within 1 min, and the

Peptide	Peak No.	m/z	position
10	Trypsin-1	2852.5 (calcd. 2852.3 for $(M+H)^+$)	8–16 95–106 111–113
	Trypsin-2	4313.7 (calcd. 4313.9 for $(M+H)^+$, average)	17–33 34–44 114–121
15	Trypsin-1′	3055.7 (calcd. 3055.3 for $(M+H)^+$)	8–16 95–106 111–113
	Trypsin-2′	4313.8 (calcd. 4313.9 for $(M+H)^+$, average)	17–33 34–44 114–121
10	GluC-1	1971.5 (calcd. 1970.9 for $(M+H)^+$)	8–16 102–106 111–113
15	GluC-1′	2174.9 (calcd. 2174.0 for $(M+H)^+$)	8–16 102–106 111–113

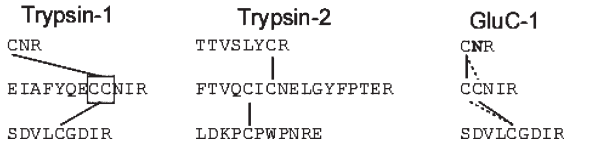
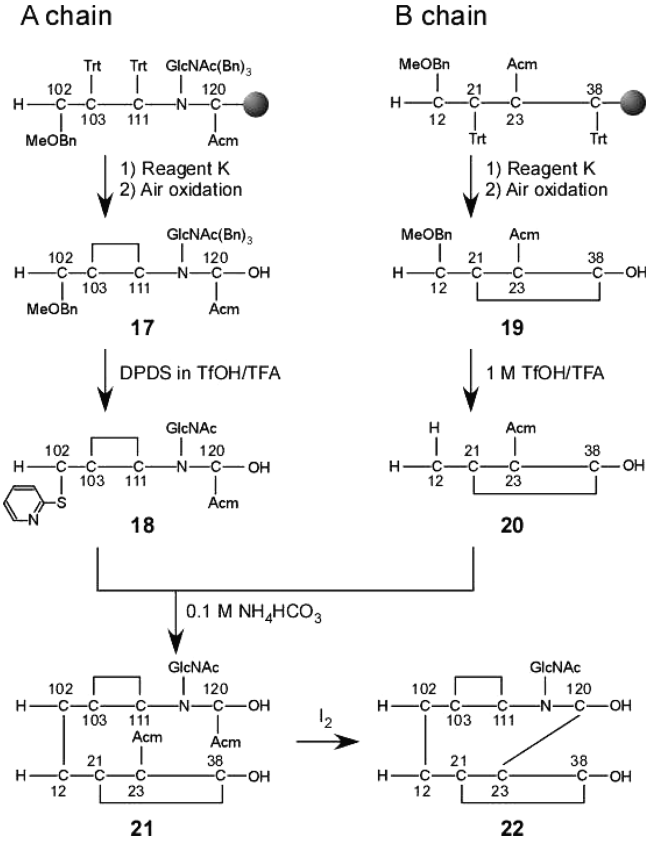


FIGURE 7: Analytical data and structures of protease digests. Solid lines indicate the disulfide bonds. Dotted lines indicate the disulfide bonds observed in the recombinant AGH expressed in a baculovirus expression system.

Scheme 3: Synthesis Procedure for $Asn(GlcNAc)^{112}$-AGH **22**



solution was kept at room temperature. The reaction was almost completed within 2 min and was stopped by adding acetic acid. Peptide **21** was obtained by RP-HPLC in 60% yield.

The final disulfide bond was formed by iodine oxidation. Peptide **21** was dissolved in 20% aqueous MeOH and added dropwise into the I_2 solution in MeOH containing HCl. After 25 min, the reaction was quenched with ascorbic acid, and the product was purified by ion-exchange chromatography, followed by RP-HPLC to give the final product **22** in 39% yield. The MALDI-TOF mass spectrum and amino acid composition of peptide **22** corresponded well to the theoretical values. Nonglycosylated AGH **23** was also synthesized using the same method as **22**.

CD Spectra and Biological Activities of Synthetic AGHs. The CD spectra of **22** and **23** were measured in phosphate buffer at room temperature (Figure 8). The spectrum of **22** was quite similar to that of **15**, demonstrating that the disulfide swapping did not significantly affect the overall conformation. In contrast, the spectral pattern of **23** differed slightly from that of **22**, suggesting that the N-glycosylation at Asn^{I12} affected the conformation.

The biological activities of **22** and **23** were assessed by *in vivo* bioassay. Injection of monosaccharyl AGH **22** caused the masculinization of female animals at a dose of 200 ng/individual (three of ten individuals were masculinized), although nonglycosylated peptide **23** showed no biological activity, even at a dose of 1 μg/individual.

DISCUSSION

It is widely believed that a refolding reaction will transform a protein into its most thermodynamically stable form (*25*). In our AGH folding experiments, the semisynthetic AGH folded into a non-native form with wrong disulfide linkages regardless of its N-linked glycan structure. These results indicated that the native AGH conformation is not the most stable at least *in vitro*.

Our previous report showed that the recombinant AGH expressed in an *E. coli* expression system and refolded in a redox buffer had the same disulfide bond arrangement as that of the baculovirus recombinant AGH (*6*). However, *in vitro* refolding reaction of the semisynthetic AGH under the conditions identical to those for the recombinant AGH gave only the disulfide isomer. These results were contradictory and required reexamination of the disulfide bond pairing in the recombinant AGH.

In order to cleave the C-peptides from the semisynthetic AGHs, we examined several commercially available proteases. A metalloendopeptidase from *Grifola frondosa*, which specifically cleaves the peptide bonds at the N-terminal side of a Lys residue, gave the most favorable results, although the two residues Lys-Arg were still retained at the N-terminus of the A-chain. An additional Met residue was also attached at the N-terminus of the B-chain due to the initial residue of protein translation. When recombinant AGH was expressed in the baculovirus expression system, the C-peptide part was cleaved with lysylendopeptidase, which specifically cleaves the peptide bonds at the C-terminal side of Lys residues (*6*). Although this recombinant peptide had a Lys residue at the C-terminus of the B-chain and a part of the C-peptide (11 residues) at the N-terminus of the A-chain, it showed biological activity *in vivo* at an injected dose of 1 ng/individual (*6*). Therefore, it is unlikely that the three additional residues in the semisynthetic peptide completely abolished the biological activity. Our synthetic AGH with a glycan moiety showed biological activity; however, the semisynthetic AGHs did not, demonstrating that the native-type disulfide bond arrangement is important for AGH activity.

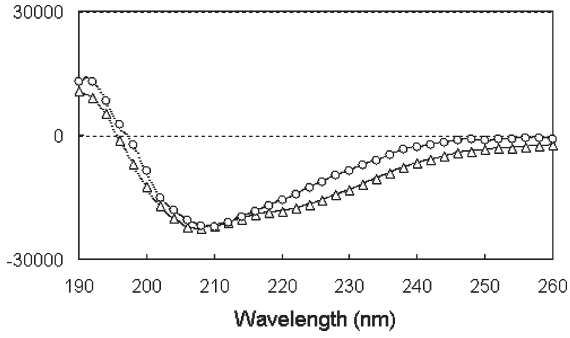


FIGURE 8: Circular dichroism spectra: circle, synthetic monosaccharyl AGH **22**; triangle, synthetic nonglycosylated AGH **23**.

Different from other insulin-family peptides, AGH has four disulfide bonds; it has an additional intrachain disulfide bond between Cys^{21}-Cys^{38} in the B-chain. In addition, the arrangement of the other three disulfide bonds is not identical to that of the other insulin-family peptides; AGH has an intrachain disulfide bond in the A-chain with 2-3 pattern rather than the traditional insulin-like 1-3 pattern. In a pioneering study, two disulfide isomers of human insulin were prepared by direct chemical synthesis (*26*). One of the isomers, which contained non-native 2-3 pattern pairings (Cys^{A7}-Cys^{A11}, Cys^{A6}-Cys^{B7}, and Cys^{A20}-Cys^{B19}), showed biological activity comparable to the native peptide (Cys^{A6}-Cys^{A11}, Cys^{A7}-Cys^{B7}, and Cys^{A20}-Cys^{B19}), indicating that a disulfide swap in insulin is not critical for its biological activity. In another previous study of bombyxin, an insulin-family peptide isolated from the silkmoth, disulfide isomers also showed biological activities comparable to the native peptide (*27*). These results differ from the present results for AGH. It has been reported that the N-linked glycan of AGH is essential for conferring AGH activity, and the results obtained in the present study support this. These observations suggest that the glycan moiety may be recognized, in part, by AGH receptor. It is likely that the disulfide swap of AGH may disrupt the formation of the folding motif which is critical for binding to the AGH receptor.

Interestingly, CD spectra of semisynthetic and synthetic AGHs were similar to those of other insulin-family peptides (*22*, *23*). The CD spectrum of a disulfide isomer of human insulin showed a lesser α-helical pattern than that of the native peptide (*23*). CD spectra of native bombyxin and its disulfide isomer showed that a disulfide swap partially disrupted the helical structure (*24*). On the other hand, the spectra of semisynthetic and synthetic monosaccharyl AGHs were similar, indicating that the disulfide bond isomerization in AGH did not significantly affect the CD spectral pattern. These results suggested that the disulfide swap in AGH did not disrupt the helical conformation typical of insulin-family peptides. Furthermore, these results also support that the AGH receptor partially recognizes the glycan moiety.

Our synthetic AGH with a GlcNAc moiety at Asn^{112} certainly possessed a biological activity *in vivo*, although it was much weaker than that of the native peptide, which showed activity at an injected dose of 38 pg/individual (*6*, *19*). It is likely that the GlcNAc moiety may not be insufficient for generating complete activity, and a larger glycan moiety may be required for full activity of AGH.

In a eukaryotic cell, polypeptide chains are synthesized on ribosomes and then transferred to the endoplasmic reticulum (ER). For glycoprotein synthesis, $Glc_3Man_9GlcNAc_2$ is first attached at an N-glycosylation site by an oligosaccharyltransferase in the ER and then rapidly trimmed by exoglycosidase to generate a $Glc_1Man_9GlcNAc_2$ polypeptide (*28*). The polypeptide portion is then folded in the ER. Glycoprotein-specific molecular chaperones, calnexin and calreticulin, are present in the ER and aid in the glycoprotein folding by binding to the glycan portion of the $Glc_1Man_9GlcNAc_2$ polypeptide (*29*). The folding pathway of AGH *in vivo* may require such molecular chaperone(s) in order to generate a thermodynamically less stable conformation.

It has been reported that human insulin-like growth factor (IGF) I is unable to form its native disulfide linkages quantitatively by *in vitro* refolding reactions, which produces not only the native conformation but also a disulfide isomer (*30*). IGF-I is coexpressed with IGF binding protein (IGFBP) I in the liver (*31*). It has been shown that IGF-I is folded quantitatively in the presence of 30 μM IGFBP-I (*32*), which suggests that IGFBP-I assists in IGF-I native disulfide formation. At present, there is no evidence that an AGH binding protein is present in the androgenic glands to direct the formation of correct (non-insulin type) disulfides in AGH.

Forming thermodynamically less stable structure is generally unfavorable for life, because more energy is required for synthesizing and maintaining the unstable structure. Why is AGH folded into the less stable form? It is likely that the degradation of the unstable protein structure is faster than that of the more stable form; i.e., the function of the less stable protein is controlled by the proteolytic pathway more easily than that of the stable form. The woodlouse may chose the such system for the strict control of sex determination.

In conclusion, we synthesized the AGH of *A. vulgare* with a homogeneous N-linked glycan moiety by the expressed protein ligation and total chemical synthesis methods. Semisynthetic AGHs had disulfide bond arrangements that differed from those of recombinant AGH prepared in the baculovirus expression system and showed no biological activities regardless of their glycan structures. In contrast, the synthetic AGH with a GlcNAc moiety showed weak biological activity. These results strongly suggest that the native conformation of AGH is not the most thermodynamically stable form and that AGH requires molecular chaperone(s) or other factor(s) for forming the correct (A-chain isomeric form of insulin) disulfide bridge arrangement. Furthermore, it was suggested that the correct disulfide linkages are necessary for conferring biological activity.

ACKNOWLEDGMENT

We are grateful to Mr. Hirotaka Haibara of the Graduate School of Agricultural and Life Sciences, The University of Tokyo, for providing the optimized conditions for the C-peptide processing reaction. We also thank Mr. Hiroaki Asai of Otsuka Chemical Co. Ltd. for providing us Fmoc-Asn[$(GalGlcNAcMan)_2$-$ManGlcNAc_2$].

REFERENCES

1. Charniaux-Cotton, H. (1954) Discovery in an amphipod crustacean (*Orchestia gammarella*) of an endocrine gland responsible for the differentiation of primary and secondary male sex characteristics. *C. R. Acad. Sci., Paris 239*, 780–782.
2. Okuno, A., Hasegawa, Y., Ohira, T., Katakura, Y., and Nagasawa, H. (1999) Characterization and cDNA cloning of androgenic gland hormone of the terrestrial isopod *Armadillidium vulgare*. *Biochem. Biophys. Res. Commun. 264*, 419–423.
3. Martin, G., Sorokine, O., Moniatte, M., Bulet, P., Hetru, C., and Dorsselaer, A. V. (1999) The structure of a glycosylated protein hormone responsible for sex determination in the isopod, *Armadillidium vulgare*. *Eur. J. Biochem. 262*, 727–736
4. Ohira, T., Hasegawa, Y., Tominaga, S., Okuno, A., and Nagasawa, H. (2003) Molecular cloning and expression analysis of cDNAs encoding androgenic gland hormone precursors from two porcellionidae species, *Porcellio scaber and P. dilatatus*. *Zoolog. Sci. 20*, 75–81
5. Manor, R., Weil, S., Oren, S., Glazer, L., Aflalo, E. D., Ventura, T., Chalifa-Caspi, V., Lapidot, M., and Sagi, A. (2007) Insulin and gender: An insulin-like gene expressed exclusively in the androgenic gland of the male crayfish. *Gen. Comp. Endrocrinol. 150*, 326–336
6. Okuno, A., Hasegawa, Y., Nishiyama, M., Ohira, T., Ko, R., Kurihara, M., Matsumoto, S., and Nagasawa, H. (2002) Preparation of an active recombinant peptide of crustacean androgenic gland hormone. *Peptides 23*, 567–572
7. Muir, T. W., Sondhi, D., and Cole, P. A. (1998) Expressed protein ligation: A general method for protein engineering. *Proc. Natl. Acad. Sci. U.S.A. 95*, 6705–6710
8. Muir, T. W. (2003) Semisynthesis of proteins by expressed protein ligation. *Annu. Rev. Biochem. 72*, 249–289. Muralidharan, V., and Muir, T. W. (2006) Protein ligation: An enabling technology for the biophysical analysis of proteins. *Nat. Methods 3*, 429–438
9. Magne, A, Gerard, D., Hirth, L., and Laustriat, G. (1977) Fluorescence study of tobacco mosaic virus protein. *Biochim. Biophys. Acta 495*, 189–194
10. Liu, J., and Gin, D. Y. (2002) *C2*-amidoglycosylation. Scope and mechanism of nitrogen transfer. *J. Am. Chem. Soc. 124*, 9789–9797
11. Matsuo, I., Nakahara, Y., Ito, Y., Nukada, T., Nakahara, Y., and Ogawa, T. (1995) Synthesis of a glycopeptide carrying a N-linked core pentasaccharide. *Bioorg. Med. Chem. 3*, 1455–1463
12. King, D. S., Fields, C. G., and Fields, G. B. (1990) A cleavage method which minimizes side reactions following Fmoc sold phase peptide synthesis. *Int. J. Pept. Protein Res. 36*, 255–266
13. Tam, J. P., Heath, W. F., and Merrifield, R. B. (1986) Mechanisms for the removal of benzyl protecting groups in synthetic peptides by trifluoromethanesulfonic acid-trifluoroacetic acid-dimethyl sulfide. *J. Am. Chem. Soc. 108*, 5242–5251
14. Hojo, H., Haginoya, E., Matsumoto, Y., Nakahara, Y., Nabeshima, K., Toole, B. P., and Watanabe, Y. (2003) The first synthesis of peptide thioester carrying *N*-linked core pentasaccharide through modified Fmoc thioester preparation: Synthesis of an *N*-glycosylated Ig domain of emmprin. *Tetrahedron Lett. 44*, 2961–2964
15. Takano, Y., Hojo, H., Kojima, N., and Nakahara, Y. (2004) Synthesis of a mimic for the heterogenous surface of core 2 sialoglycan-linked glycoprotein. *Org. Lett. 6*, 3135–3138
16. Hojo, H., Matsumoto, Y., Nakahara, Y., Ito, E., Suzuki, Y., Suzuki, M., Suzuki, A., and Nakahara, Y. (2005) Chemical synthesis of 23 kDa glycoprotein by repetitive segment condensation: A synthesis of MUC2 basal motif carrying multiple O-GalNAc moieties. *J. Am. Chem. Soc. 127*, 13720–13725
17. Ye, Y. H., Li, H., and Jiang, X. (2005) DEPBT as an efficient coupling reagent for amide bond formation with remarkable resistance to racemization. *Biopolymers 80*, 172–178.
18. Yamamoto, N., Takayanagi, A., Sakakibara, T., Dawson, P. E., and Kajihara, Y. (2006) Highly efficient synthesis of sialylglycopeptides overcoming unexpected aspartimide formation during activation of Fmoc-Asn(undecadisialyloligosaccharide)-OH. *Tetrahedron Lett. 47*, 1341–1346.
19. Okuno, A., Hasegawa, Y., and Nagasawa, H. (1997) Purification and properties of androgenic gland hormone from the terrestrial isopod, *Armadillidium vulgare*. *Zoolog. Sci. 14*, 837–842.
20. Hasegawa, Y., Haino-Fukushima, K., and Katakura, Y. (1987) Isolation and properties of androgenic gland hormone from the terrestrial isopod, *Armadillidium vulgare*. *Gen. Comp. Endrocrinol. 67*, 101–110.
21. Dawson, P. E., Muir, T. W., Clark-Lewis, I., and Kent, S. B. (1994) Synthesis of proteins by native chemical ligation. *Science 266*, 776–779.
22. Hua, Q.-X., Jia, W., Frank, B. H., Phillips, N. F. B., and Weiss, M. A. (2002) A protein caught in a kinetic trap: Structures and stabilities of insulin disulfide isomers. *Biochemistry 41*, 14700–14715.
23. Hua, Q.-X., Mayer, J. P., Jia, W., Zhang, J., and Weiss, M. A. (2006) The folding nucleus of the insulin superfamily: A flexible peptide model foreshadows the native state. *J. Biol. Chem. 281*, 28131–28142.
24. Maruyama, K., Nagata, K., Tanaka, M., Nagasawa, H., Isogai, A., Ishizaki, H., and Suzuki, A. (1992) Synthesis of bombyxin-IV, and insulin superfamily peptide from the silkworm, *Bombyx mori*, by

stepwise and selective formation of three disulfide bridges. *J. Protein Chem. 11*, 1–12.
25. Anfinsen, C. B. (1973) Principles that govern the folding of protein chains. *Science 181*, 223–230.
26. Sieber, P. S., Eisler, K., Kamber, B., Riniker, B., Rittel, W., Marki, F., and deGasparo, M. (1978) Synthesis and biological activity of two disulphide bond isomer of human insulin [A7-A11,A6-B7-cystine]- and [A6-A7,A11-B7-cystine]insulin (human). *Hoppe-Seyler's Z. Physiol. Chem. 359*, 113–123.
27. Nagata, K., Maruyama, K., Nagasawa, H., Urushibata, I., Isogai, A., Ishizaki, H., and Suzuki, A. (1992) Bombyxin-II and its disulfide bond isomers: Synthesis and activity. *Peptides 13*, 653–662.
28. Kornfeld, R., and Kornfeld, S. (1985) Assembly of asparagines-linked oligosaccharides. *Annu. Rev. Biochem. 54*, 631–664.
29. Williams, D. B. (2006) Beyond lectins: The calnexin/calreticulin chaperone system of the endoplasmic reticulum. *J. Cell Sci. 119*, 615–623. Ellgaard, L., and Frickel, E. M. (2003) Calnexin, calreticulin, and ERp57: Teammates in glycoprotein folding. *Cell Biochem. Biophys. 39*, 223–247.
30. Hober, S., Forsberg, G., Palm, G., Hartmanis, M., and Nilsson, B. (1992) Disulfide exchange folding of insulin-like growth factor I. *Biochemistry 31*, 1749–1756.
31. Humbel, R. E. (1990) Insulin-like growth factors I and II. *Eur. J. Biochem. 190*, 445–462.
32. Hober, S., Hansson, A., Uhlen, M., and Nilsson, B. (1994) Folding of insulin-like growth factor I is thermodynamically controlled by insulin-like growth factor binding protein. *Biochemistry 33*, 6758–6761.