Evidence for Unique Homologous Peptide Sequences around the Glycosylated Seryl and Threonyl Residues in Polysialoglycoproteins Isolated from the Unfertilized Eggs of the Pacific Salmon *Oncorhynchus keta*

Michio Shimamura and Yasuo Inoue*

Department of Biophysics and Biochemistry, Faculty of Science, University of Tokyo, Hongo-7, Tokyo 113, Japan

Sadako Inoue

School of Pharmaceutical Sciences, Showa University, Hatanodai-1, Tokyo 142, Japan

Received December 10, 1984

ABSTRACT: Structures of glycopeptides obtained by exhaustive Pronase digestion of high molecular weight (1.7×10^5) salmon egg polysialoglycoprotein have been elucidated. Six principal glycopeptides isolated by gel chromatography and DEAE-Sephadex A-25 chromatography in the absence or presence of borate ion were analyzed for their carbohydrate and amino acid composition, as well as amino acid sequence, and found to be of two distinct types: (1) glycotripeptides, Thr*-Ser*-Glu, and (2) glycotetrapeptides, Thr*-Gly-Pro-Ser, where an asterisk indicates the amino acid residues to which either the Galβ1→3GalNAc or Fucα1→3GalNAcβ1→3Galβ1→4Galβ1→3GalNAc chain is attached. Their final yield corresponds to 64% of the original desialylated glycoprotein. In view of the simple amino acid composition of salmon egg polysialoglycoprotein (molar ratio $Asp_2Thr_2Ser_3Glu_1Pro_1Gly_1Ala_3$) and the result of alkaline β-elimination indicating three carbohydrate units linked to two of two threonine and one of three serine residues, a unique primary structure comprising repetitive sequences of the above two types of glycopeptides, which are interspersed by short nonglycosylated peptides consisting of alanine and aspartic acid, has been proposed for the core protein. The molecular secondary ion mass spectra of underivatized glycopeptides were used to obtain their structural information. The anomeric configuration of the proximal sugar–peptide linkages was proven to be α by proton nuclear magnetic resonance spectroscopy. This is the first systematic reported study of O-glycosidically linked glycopeptides by these instrumental methods.

Since the first demonstration of the existence of polysialoglycoprotein (PSGP)[1] in rainbow trout eggs in 1978 (Inoue & Iwasaki, 1978), it has become apparent that this novel class of glycoprotein is a prominent constituent of the mature oocytes of all species of *Salmonidae* studied (*Oncorhynchus keta*, *Oncorhynchus masou*, *Oncorhynchus nerka*, Salmo *gairdneri irideus*, and *Salvelinus leucomaenis pluvius*) (S. Inoue and M. Iwasaki, unpublished results). The amino acid compositions of PSGP from these sources are very similar. PSGPs isolated from unfertilized, unactivated eggs are of high molecular weight ($\sim 2 \times 10^5$) and polydisperse. The main structural features of PSGP are as follows: these molecules are made up of a central protein core ($\sim$15% by weight) to which chains of oligo- and/or polysialylglycan ($\sim$85%) are attached through O-glycosidic linkages (Inoue & Iwasaki, 1980; Nomoto et al., 1982; Shimamura et al., 1983, 1984a; Iwasaki et al., 1984a,b). The most striking feature of PSGP is the high content (60% or more by weight) of sialic acid occurring mostly in the form of [→8NeuGcα2→]$_n$. Although the biological functions of PSGP have not yet been established, their high content of sialic acid residues and relatively high proportion of acidic amino acid residues ($\sim$20%) would suggest a possible role in regulation of cations such as H^+ and Na^+ and/or transport of Ca^{2+} in the egg cells at fertilization and subsequent early development (Gilkey, 1981).

Any studies concerning the functional significance of PSGP will require knowledge of the structure of both the prosthetic carbohydrate units and the core protein. This investigation was initiated as a part of a program directed toward the elucidation of the complete structure of PSGP. Whereas a considerable amount is known about the structure of the carbohydrate chains (Inoue & Matsumura, 1979; Inoue & Iwasaki, 1980; Inoue et al., 1981; Nomoto et al., 1982; Iwasaki et al., 1984a,b; Kitajima et al., 1984; Shimamura et al., 1983,1984a), there is little information on the peptide part of the molecule: In the case of the eggs of Pacific salmon, about 85% of the dry weight of the PSGP is accounted for by carbohydrates, mostly Galβ1→3[(→8NeuGcα2)$_n$→6]GalNAc and Fucα1→3GalNAcβ1→3Galβ1→4Galβ1→3[(→8NeuGcα2)$_n$→6]GalNAc attached O-glycosidically to the hydroxyl groups of Ser and Thr residues of the core protein (Shimamura et al., 1984a). The presence of minor carbohydrate units, GalNAcβ1→4[NeuGcα2→3]GalNAcβ1→3Galβ1→4Galβ1→3[(→8NeuGcα2)$_n$→6]GalNAc, has also been established (Shimamura et al., 1984a).

The results given in this paper are concerned with structural information on the sequence of the amino acids in the region around the points of attachment of these carbohydrates to the protein. We decided to assess the effect of Pronase on the purified salmon egg PSGP and the desialylated PSGP. Chromatography of the digestion products of the desialylated PSGP led to the isolation of six major glycopeptides with a peptide length of three and four amino acid residues. Glycopeptides thus obtained were shown to be of two distinct types, Thr*-Ser*-Glu and Thr*-Gly-Pro-Ser, in which (*) indicates the amino acid residues to which either asialo short-core unit

[1] Abbreviations: PSGP, polysialoglycoprotein; Fuc, L-fucose; NeuGc, *N*-glycolylneuraminic acid; GalNAcol, *N*-acetyl-D-galactosaminitol; Tris, tris(hydroxymethyl)aminomethane; dansyl, 5-(dimethylamino)-naphthalene-1-sulfonate; NMR, nuclear magnetic resonance; SIMS, secondary ion mass spectrometry.

0006-2960/85/0424-5470$01.50/0

or asialo Fuc-containing unit is attached. The seryl residue adjacent and C-terminal to Pro was found to be devoid of carbohydrate. Thus, the carbohydrate chains on PSGP are localized to specific Thr and Ser residues of the molecule, with most, if not all, oligo- and/or polysialylglycan units confined to the above two types of glycotri- and glycotetrapeptides. The present study also revealed that the primary structure of the salmon egg desialylated PSGP is built of internally repeating sequences of the above glycotri- and glycotetrapeptides, which are interspersed by rather short non-glycosylated peptides accessible to Pronase.

Materials and Methods

Desialylation of PSGP. PSGP was isolated and purified from the unfertilized eggs of *Oncorhynchus keta* as previously described (Shimamura et al., 1983). Removal of sialic acid from PSGP was performed first with dilute acid and then with sialidase. PSGP (423 mg) was dissolved in 30 mL of dilute HCl (pH 2.0) and heated for 30 min at 80 °C under a nitrogen atmosphere. After cooling and neutralizing, the solution was applied to a Sephadex G-75 column (2.1 × 135 cm, pre-equilibrated with 50 mM NH_4HCO_3), and the excluded glycoprotein material was pooled and evaporated to remove excess salt. Partially desialylated PSGP, dissolved in 20 mL of 0.1 M sodium acetate buffer (pH 4.7), was subjected to exhaustive digestion with *Arthrobacter ureafaciens* sialidase by stepwise addition (total 6 times) of a total of 5.0 units under toluene for 167 h at 27 °C. These treatments afforded a desialylated PSGP in which a small amount (~1%) of NeuGc was still detectable. Most of these residues can be accounted for by the presence of sialidase-resistant NeuGc in the long-core units found in the salmon egg PSGP (Shimamura et al., 1984a).

Pronase Digestion of Desialylated PSGP. Pronase digestion of the desialylated PSGP followed closely the procedure described by Spiro (1966): Desialylated PSGP was dissolved in 10 mL of 1.5 mM $CaCl_2$ in 0.1 M Tris–acetic acid buffer (pH 7.8) and treated with Pronase P (Kaken Kagaku, Tokyo; 750 000 Tyr units/g); the mixture was incubated for a total of 6 days under a toluene atmosphere at 37 °C. The enzyme (total 5 mg) was added in four proportions. At every addition of Pronase, the pH was adjusted to 7.8 with Tris.

Isolation and Purification of Glycopeptides. The digest was first fractionated by gel filtration on a column (1.6 × 145 cm) of Bio Gel P-4 (200–400 mesh) equilibrated with 0.1 M pyridine–acetic acid, pH 5.0. The column was eluted with the same buffer at 7 mL/h, and 4.5-mL fractions were collected. Fractions were monitored by the ninhydrin reaction (Moore & Stein, 1954) and by phenol–sulfuric acid (Dubois et al., 1956). The Bio Gel P-4 column was calibrated with glycine and triglycine. Blue dextran was used to determine the void volume, V_0. Fractions GP1 to GP4 in Figure 1A were combined, concentrated, desalted, and redigested in 2 mL of 0.1 M Tris–acetic acid/1.5 mM $CaCl_2$, pH 7.7, with Pronase (in four proportions of a total of 1 mg) at 37 °C under toluene for 5 days. The digest was reapplied to the Bio Gel P-4 column under identical conditions as in Figure 1A. The glycopeptide-containing fractions (GP1 to GP6) were examined by high-voltage paper electrophoresis. The electrophoresis was performed in 10 mM KH_2PO_4–Na_2HPO_4 buffer, pH 8.0, at 45 V/cm for 1.5 h with Toyo-roshi No. 51A paper, and the peptide spots were detected by spraying with 2% ninhydrin reagent in acetone and heating at 100 °C for 5 min. The phenol–sulfuric acid positive peaks GP1 through GP6 were further fractionated by DEAE-Sephadex A-25 column chromatography in the presence or absence of borate buffer (Nomoto & Inoue, 1983).

Polyacrylamide Gel Electrophoresis. Electrophoresis in a single dimension was performed according to the procedure of Laemmli (1970). Following the run, gels were fixed and then stained with periodic acid–Schiff reagent.

Ultracentrifugal Analysis. The sample dissolved at 0.5% in 0.5 M NaCl buffered with 0.01 M Tris-HCl (pH 8.0) was subjected to sedimentation-equilibrium analysis using a multichannel cell (Yphantis, 1960) on a Hitachi ultracentrifuge UCA-1A. Apparent partial specific volume was calculated to be 0.57 from the composition of PSGP.

Analysis of Carbohydrate Contents. Carbohydrates were determined by gas–liquid chromatography by the method of Reinhold (1972) after methanolysis with 0.5 M methanolic HCl followed by N-acetylation (Kozulič et al., 1979) and trimethylsilylation. The trimethylsilylated derivatives of the methyl glycosides were separated on a glass column (3 mm × 100 cm) packed with Chromosorb W coated with 1.5% OV-17. A Shimadzu GC-4BM gas chromatograph with flame ionization detector was programmed for a temperature gradient from 160 to 240 °C at 4 °C/min. Peak areas were measured with a Shimadzu Chromatopac C-R1A. The hexose content was estimated by the phenol–sulfuric acid method (Dubois et al., 1956). Galactosamine was determined on an amino acid analyzer: The glycopeptides were hydrolyzed in 4 N HCl for 4 h at 100 °C.

Amino Acid Analysis. PSGP and glycopeptides were analyzed for amino acid content with a Hitachi Model KLA-5 amino acid analyzer after hydrolysis in vacuo in 5.7 N constant-boiling HCl at 110 °C for 4 h. A hydrolysis time of 4 h gave the maximum yield of each component amino acid (Shimamura et al., 1984b).

Alkali and Alkaline Borohydride Treatments. Alkali lability of the sugar–peptide linkages was examined by incubating glycopeptides (0.1 μmol) either with 0.1 N NaOH or with 1 M $NaBH_4$ in 0.1 N NaOH at 37 °C for 48 h. Reduction was stopped by acidification to pH 5.0 with dilute acetic acid, and excess $NaBH_4$ was removed as methyl borate.

Amino-Terminal Analysis. This was carried out by the dansylation procedure of Gray (1972a). Dansylated amino acids were separated and identified by chromatography on Kiesel gel 60 thin-layer plates with the solvent system (Morse & Horecker, 1966) $CHCl_3$/2-methyl-2-propanol/acetic acid (70:30:3 by volume) or ethyl acetate/methanol/acetic acid (20:1:1 by volume).

Manual Dansyl Edman Degradation. Amino acid sequencing of glycopeptides was performed by the method of Gray (1972b) starting with 0.3–0.4 μmol of each glycopeptide. For GP4-4, GP3-1, and GP2-5, 50 nmol of the sample was subjected to Edman degradation by the method of Tomita et al. (1978).

Carboxyl-Terminal Amino Acid Analysis. The carboxyl-terminal sequence was established or confirmed by digestion with carboxypeptidase Y (Oriental Yeast, Tokyo) according to the procedure described by Hayashi et al. (1975). For this experiment, 0.1 μmol each of GP3-1 and GP4-4 was dissolved in 0.2 mL of 50 mM sodium phosphate buffer (pH 6.5) and incubated with 4 μg (or 0.04 unit) of CPase Y at 37 °C for 1 h. The liberated amino acid(s) was (were) identified by the dansyl method of Gray (1972a), using thin-layer chromatography on Kiesel gel plates.

Subtilisin Degradation. Subtilisin Carlsberg (Sigma Chemical Co.) treatment of GP5-1 and GP6-1, 0.25 μmol each, was performed in 0.1 mL of 0.1 M ammonium acetate (pH 8.0) by addition in two proportions of the enzyme (total 20 μg, 0.24 unit) (Bennett, 1967). The mixture was incubated

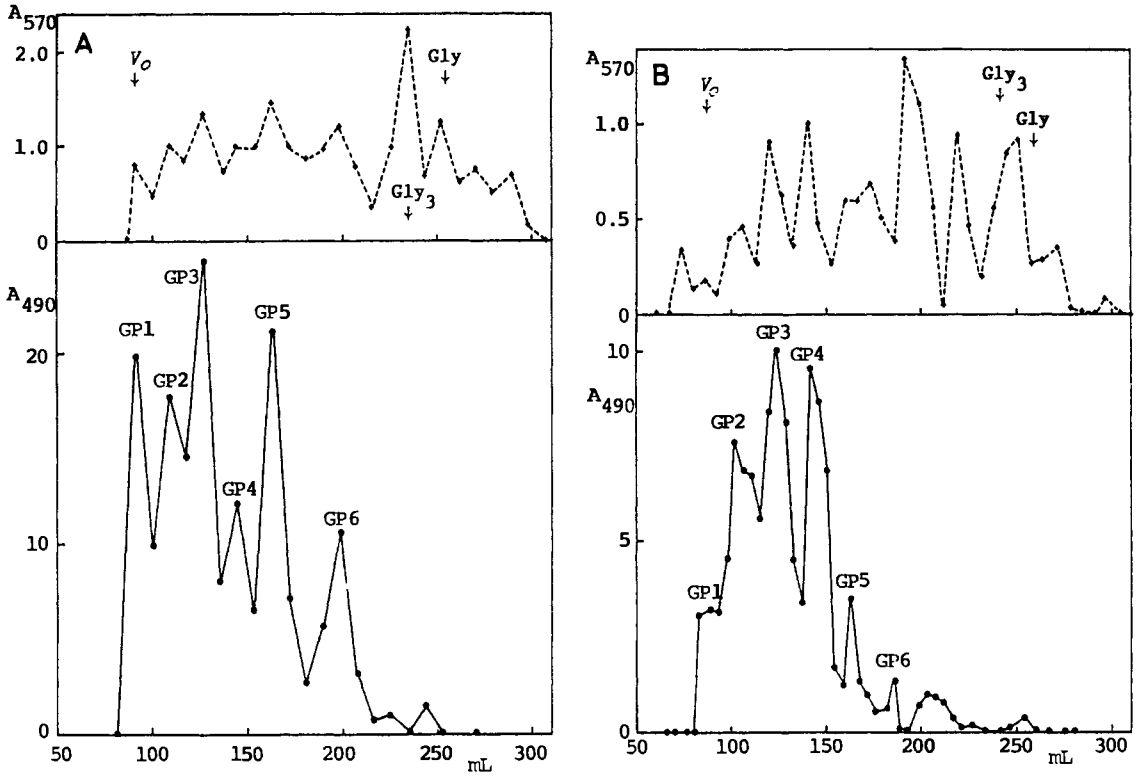


FIGURE 1: (A) Gel filtration of the Pronase digest of desialylated salmon egg PSGP. The Pronase-digested peptides were subjected to chromatography on a column (1.6 × 145 cm) of Bio-Gel P-4 equilibrated with 0.1 M pyridine–acetic acid, pH 5.0, at flow rate of 7 mL/h. Fractions of 4.5 mL were collected and assayed for hexose (—) and peptide material (--) by the phenol–sulfuric acid method and a manual ninhydrin method, respectively. The fractions under the peaks indicated were pooled to give GP5 and GP6. (B) The higher molecular weight glycopeptide fractions GP1–GP4 were pooled and further digested with Pronase. The products of this digestion were rechromatographed on the same column, with the profiles aligned with that shown in (A) by alignment of the peak positions of the three internal standards, Gly, Gly_3, and blue dextran. Arabic numerals in GP1–GP6 correspond to glycopeptide products of decreasing molecular size.

at 37 °C for 25 h. The released amino acids were dansylated by the procedure of Gray (1972a) and identified by chromatography on silica gel plates.

Secondary Ion Mass Spectrometry. Positive-ion SIMS spectra of the glycopeptides with molecular weight below 1500 were obtained in a manner similar to that described in previous papers (Shimamura et al., 1983, 1984a) with a Hitachi M-80 double-focusing mass spectrometer equipped with a sputtered-ion MS system and an M-003 data processor at Central Research Laboratory, Hitachi, Ltd.

^1H NMR Spectroscopy. The 270-MHz ^1H NMR spectra were recorded on a Burker WH-270 spectrometer at 23 and 50 °C. Chemical shifts are given in ppm relative to that of sodium 4,4-dimethyl-4-silapentane-1-sulfonate (DSS) and indirectly to acetone (δ 2.225 ppm at 23 °C) in D_2O. The probable errors for chemical shift (δ) and coupling constant (J) values are estimated to be ±0.001 ppm and ±0.3 Hz, respectively, under the experimental conditions employed.

Results

Salmon egg PSGP showed rather broad molecular weight dispersion when examined by gel filtration chromatography and polyacrylamide gel electrophoresis (results not shown). A high molecular weight fraction slightly retained by a Sephadex G-200 column in 0.1 M NaCl has been used in this and previous structural studies. The weight-average molecular weight of this fraction was determined to be 1.7×10^5 by sedimentation-equilibrium analysis. This fraction of PSGP gave a diffuse band on polyacrylamide gel electrophoresis (10% acrylamide) though it scarcely entered the gel. The PSGP stained with periodic acid–Schiff reagent but not with Coomassie blue. Other protein constituents were not detected by staining with either periodic acid–Schiff reagent or Coomassie blue. The analytical data for salmon egg PSGP used as the starting material are given in the last column of Table I. Only aspartic acid was detected as the amino-terminal amino acid, thus confirming the homogeneity of the PSGP preparation.

Pronase Digestion. Pronase acted more extensively on desialylated PSGP than the intact PSGP, and for this reason, desialylated material was used in the present work. The elution profile from a Bio-Gel P-4 column of the Pronase digest of desialylated PSGP was obtained by ninhydrin and phenol–sulfuric acid assays and showed six discrete glycopeptide peaks labeled GP1–GP6 (Figure 1A). In order to examine possible incomplete digestion of the larger molecular weight fractions, materials under peaks GP1–GP4 in Figure 1A were pooled, evaporated, and redigested with Pronase. The gel filtration procedure was repeated as before (Figure 1B). The elution profile of the redigested fractions was only slightly changed, and this indicates that, after two rounds of Pronase treatment, digestion was essentially complete and that the fractions GP1–GP6 represent different molecular entities on the basis of their size.

Fractionation of Glycopeptides GP1–GP6 according to Glycosylation Pattern. For fractions GP1–GP6, a homogeneity test was carried out by high-voltage paper electrophoresis. At pH 8.0, all the fractions except GP1 showed more than one ninhydrin-positive spot. The relatively low molecular weight glycopeptides, GP3–GP6, were further purified by anion-exchange chromatography on DEAE-Sephadex A-25 either in the absence or in the presence of borate ions (Figure 2). The elution profile of GP6 (Figure 2A) showed three peaks when monitored by absorbance at 220 nm, but GP6-1 was found to be the only carbohydrate-containing peak that was neutral

Table I: Chemical Compositions of Pronase-Derived Glycopeptides of Desialylated PSGP[a]

	GP6-1	GP5-1	GP5-2	GP4-3	GP4-4	GP3-1	PSGP
Asp			0.30 (0)[b]	1.69		0.14 (0)[b]	2.1[c]
Thr	0.98 (1)[b]	1.03 (1)[b]	1.07 (1)	1.93	0.98 (1)[b]	1.21 (1)	1.6
Ser	0.99 (1)	0.94 (1)	1.07 (1)	1.26	0.90 (1)	1.16 (1)	2.8
Glu			1.00 (1)	0.80	1.00 (1)	1.00 (1)	0.9
Pro	1.00 (1)	1.00 (1)	0.21 (0)	1.00		0.13 (0)	1.0
Gly	1.00 (1)	1.08 (1)	0.38 (0)	1.83		0.29 (0)	1.5
Ala				0.56		0.20 (0)	2.4
GalNAc	0.95 (1)	2.04 (2)	1.73 (2)	1.5	2.57 (3)	3.60 (4)	3.6
Gal	1.1 (1)	2.0 (2)	1.9 (2)	2.4	2.7 (3)	4.6 (4)	5.2
Fuc		1.0 (1)		1.3	0.7 (1)	1.8 (2)	1.5
NeuGc							18
yield (%)	22	6	17	(11)	11	8	

[a] Amino acids for which <0.1 mol/mol was determined are omitted. [b] Expected values are in parentheses, based on the structures shown in Figure 3. [c] Values are molar ratios relative to Pro taken as 1.0.

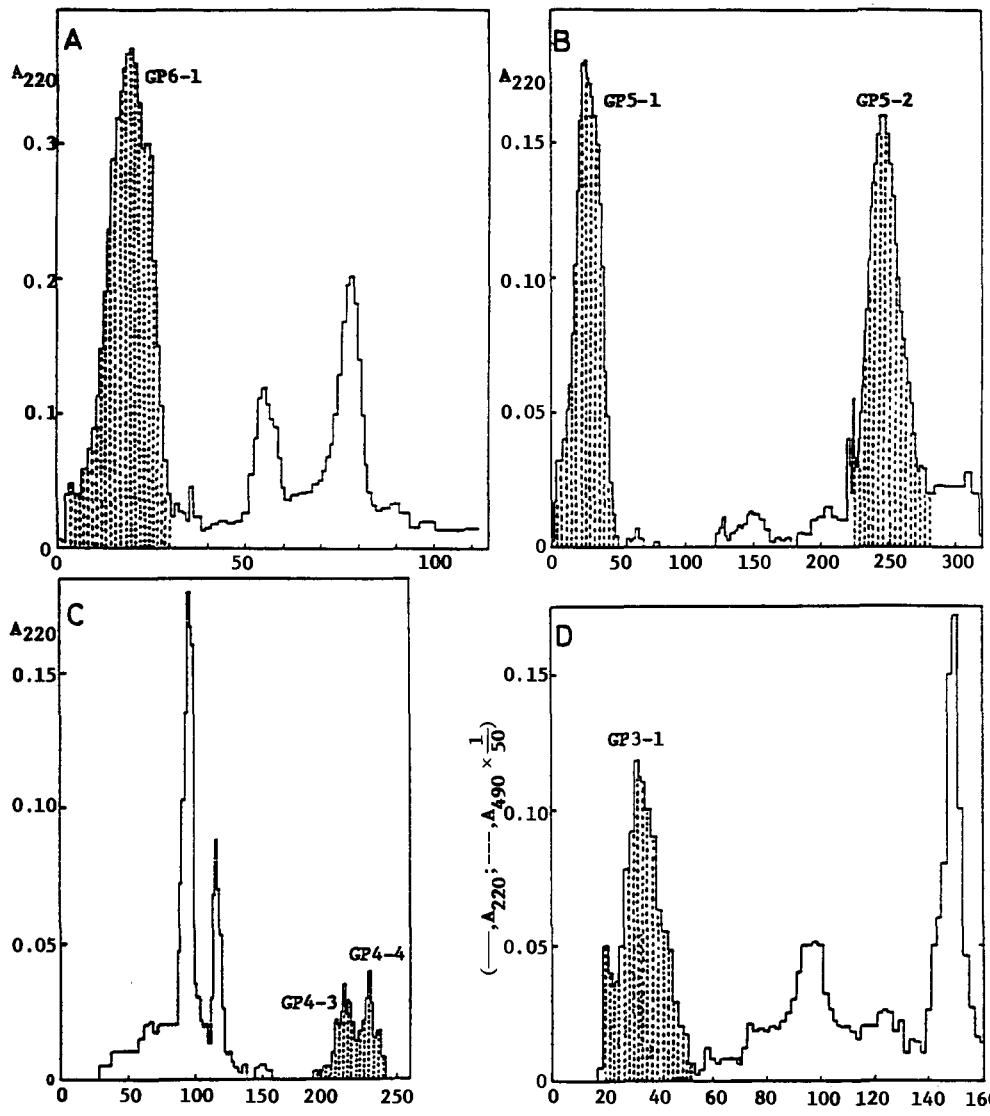


FIGURE 2: Separation of glycopeptides by ion-exchange chromatography on DEAE-Sephadex A-25 (Cl^- form unless otherwise specified) column: (A) GP6, obtained as shown in Figure 1A, was applied to a column 0.8 × 58 cm. The column was eluted with a linear gradient of NaCl in 4 mM Tris-HCl (pH 8.0) from 0 (300 mL) to 0.15 M (300 mL); the fraction volume was 4.9 mL; the flow rate was 12 mL/h. (B) GP5 was applied to the column (0.8 × 56 cm). Elution was effeced by a linear gradient of NaCl in 4 mM Tris-HCl (pH 8.5) from 0 (500 mL) to 0.05 M (500 mL). (C) GP4 was applied to a column (1.2 × 83 cm, borate ion form), and the column was developed with a linear gradient of sodium borate (pH 8.0) from 0.01 to 0.6 M and then with 0.2 M NaCl after fraction 200. (D) GP3 on a column (0.8 × 68 cm) with a linear gradient of NaCl in 5 mM Tris-HCl, pH 7.6, from 0 (300 mL) to 0.15 M (300 mL). Phenol–sulfuric acid positive peaks are indicated by stippling. The abscissas express fraction number.

on paper electrophoresis at pH 8.0.

GP5, which was eluted in the position of molecular size slightly larger than M_r 700 on gel chromatography, gave two phenol–sulfuric acid positive peaks (Figure 2B). In view of their migrating behavior on paper electrophoresis and elution positions in DEAE-Sephadex A-25 chromatography, these components, GP5-1 and GP5-2, were presumed to be the neutral and monoanionic glycopeptides, respectively. Fraction

Table II: Amino Acid Composition of Glycopeptides GP6-1, GP5-1, GP5-2, and GP4-4 after Treatment with 0.1 M NaOH/1 M $NaBH_4$ at 37 °C for 48 h[a]

	GP6-1	GP5-1	GP5-2	GP4-4	GP6-1[c]
Asp			0.38 (+0.08)		
Thr	0.89 (−0.11)	0.97 (−0.06)	0.89 (−0.18)	1.08 (+0.10)	0.85 (−0.15)
Ser	1.00 (+0.01)	0.93 (−0.01)	0.28 (−0.79)	0.19 (−0.81)	1.02 (+0.03)
Glu			1.00	1.00	
Pro	1.00	1.00			1.00
Gly	0.21 (−0.79)[b]	0.21 (−0.87)[b]	0.22 (−0.16)		0.94 (−0.06)
Ala			0.76 (+0.76)	0.61 (+0.61)	

[a] Values are expressed as molar ratio to proline for GP6-1 and GP5-1 and to glutamic acid for GP5-2 and GP4-4. [b] Refer to Shimamura et al. (1984b) for a marked loss of the glycine residue. [c] After treatment with 0.1 M NaOH at 37 °C for 48 h in the absence of BH_4^-.

GP4 was resolved into several peaks (Figure 2C) when fractionated on DEAE-Sephadex A-25 in the presence of borate buffer, but only GP4-3 and GP4-4 were phenol–sulfuric acid positive. GP3 gave one major, carbohydrate-containing peak on anion-exchange chromatography (Figure 2D).

Composition Analysis of Glycopeptides. Amino acid and carbohydrate compositions of the glycopeptides isolated from desialylated PSGP are given in Table I. The compositions of Gp6-1, GP5-1, and GP4-3 are expressed as residues per glycopeptide molecule with Pro set at 1.00, and those of GP5-2, GP4-4, and GP3-1 are presented as residues per glycopeptide with Glu set at 1.00. A pair of GP6-1 and GP5-1, and likewise GP5-2, GP4-4, and GP3-1, appears to differ only in the nature of carbohydrate chains.

The glycopeptides were treated with alkaline borohydride, and the amino acid composition of the treated glycopeptides is shown in Table II. As judged from the composition of GP6-1 and GP5-1 (Table I), these two glycopeptides appear to represent monoglycosylated materials with the carbohydrate chain on either Thr or Ser residues. The BH_4^-/OH^- treatment of GP6-1 and GP5-1 resulted in no significant loss of Thr and Ser (Table II).

This, taken together with the fact that the GalNAc was almost completely resistant to alkaline borohydride reaction, would suggest that the carbohydrate chain is at the N-terminal Thr or Ser residue (Pigman et al., 1974). Since GP5-2 and GP3-1 were found to have an amino acid composition (Table I) consistent with being composed of approximately equimolar amounts of Thr, Ser, and Glu residues, the Asp, Pro, Gly, and Ala values are probably not significant. Due to insufficient material in any of these fractions, they were subjected to structural studies without further purification.

Amino Acid Sequencing of the Glycopeptides GP6-1, GP5-1, GP5-2, GP4-4, and GP3-1. All of these glycopeptides, when subjected to the dansylation method of Gray (1972a) followed by acid hydrolysis and thin-layer chromatography, showed Thr as the only amino-terminal amino acid. First, the peptide sequence of GP6-1 was determined by manual dansyl-Edman degradation to be Thr-Gly-Pro-Ser. Alkali-resistant *O*-glycosides can usually be encountered if the substituted Ser or Thr residues are in amino- or carboxyl-terminal positions (Derevitskaya et al., 1967; Pigman et al., 1974), so that results of BH_4^-/OH^- treatment (Table II) were inconclusive to assign the glycosylated amino acid residue. When GP6-1 was digested with subtilisin followed by dansylation, no dansylthreonine could be observed whereas dansylated Gly, Pro, and Ser were recovered. It was therefore concluded that the carbohydrate (*) moiety was located at the N-terminal Thr residue as indicated in the structure for GP6-1 (see also Figure 3): Thr*-Gly-Pro-Ser where (*) is the short core Galβ1→3GalNAc, as judged by composition analysis (Table I). It was not considered necessary to reinvestigate the carbohydrate structures for which three types have already been established before (Shimamura et al., 1984a).

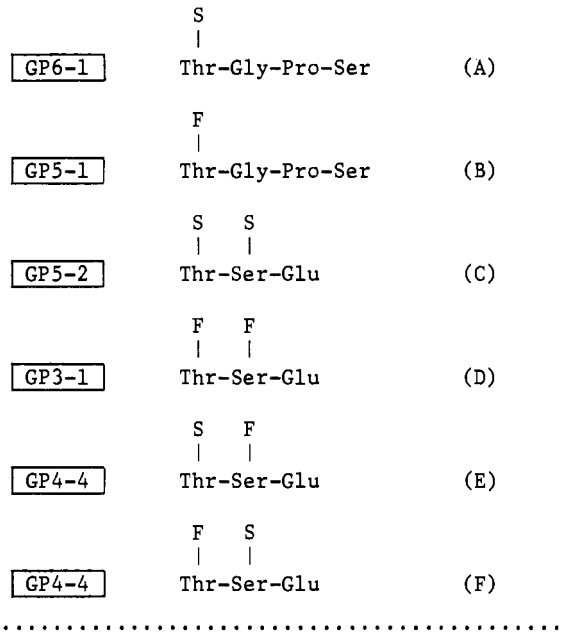


FIGURE 3: Glycopeptide structures.

The peptide sequence determined for GP5-1 was identical with that found in GP6-1. The position to which the carbohydrate chain was attached was determined as shown for GP6-1. The carbohydrate compositions differ (Table I), which may account for their resolution on Bio-Gel P-4. GP5-1 contains 2 mol of Gal and GalNAc each and 1 mol of Fuc, suggesting that a pentasaccharide, Fucα1→3GalNAcβ1→3Galβ1→4Galβ1→3GalNAc, is located at the glycosylation site (structure B in Figure 3).

When GP4-4 was subjected to paper electrophoresis, a single spot was present. The peptide sequence of GP4-4 was determined by using the manual Edman degradation procedure (first two residues) and carbocypeptidase Y digestion method (last residue): Thr-Ser-Glu. The Ser content decreased by 90% (Table II) on alkaline borohydride reaction, but about half of the total carbohydrate of the original GP4-4 was retained (not shown), indicating that this glycopeptide contains two carbohydrate chains with one on Thr and the other on Ser residues. Furthermore, the carbohydrate-composition analysis before and after alkaline borohydride treatment suggests that the two chains are nonidentical and that GP4-4 represents a diglycosylated tripeptide composed of one short core and one Fuc-containing core. The results also suggest heterogeneity in the distribution of these two types of carbohydrate chains attached to Thr and Ser residues (see ^1H NMR spectral

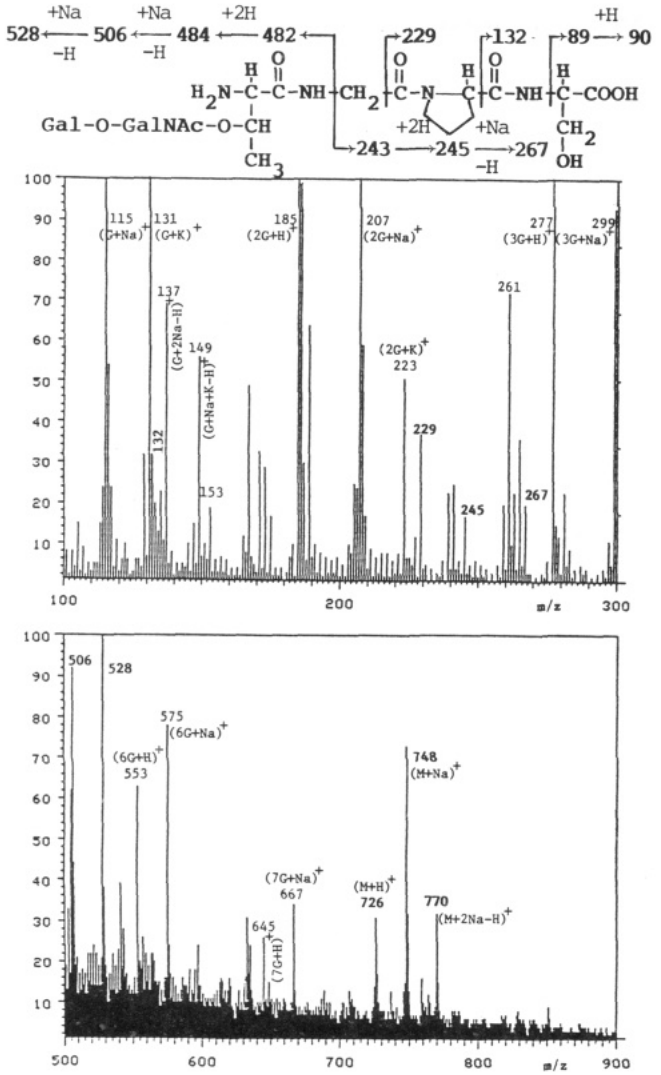


FIGURE 4: Positive-ion SIMS spectrum of GP6-1. The structure and masses of the proposed fragment ions are shown.

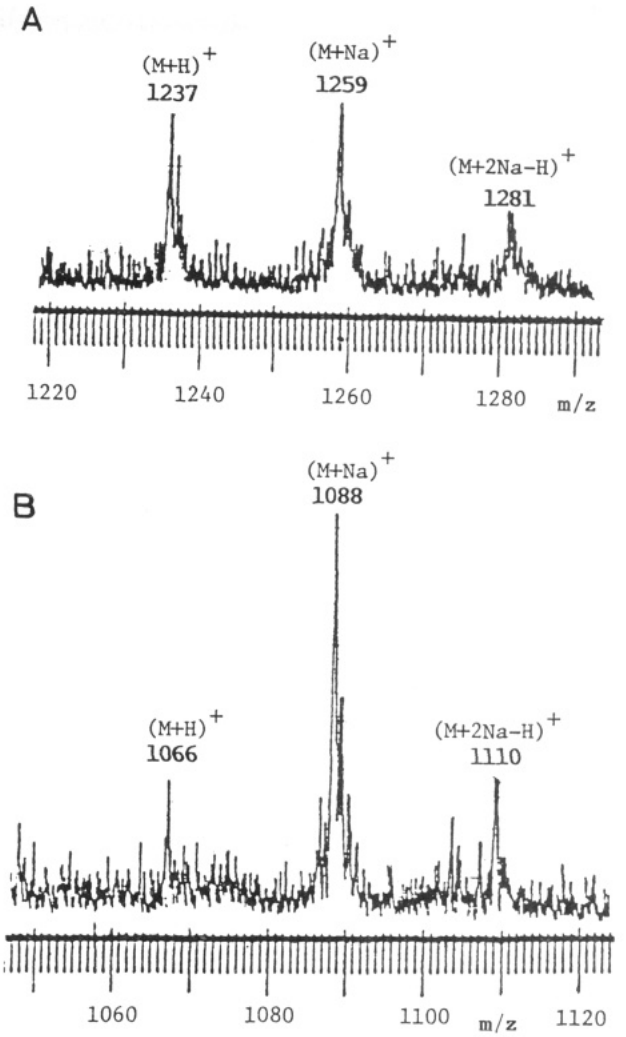


FIGURE 5: Positive-ion SIMS spectra of the intact glycopeptides (A) GP5-1 and (B) GP5-2.

analysis of GP4-4 in the following section).

Interestingly, the amino acid compositions of GP5-2 and GP3-1 were quite similar to that of GP4-4 though the fractional amounts of the amino acids other than Thr, Ser, and Glu would suggest the presence of some contaminating peptide(s), which were not separable from the major components either by virtue of similarity in molecular size or by charge. The amino acid sequence of GP5-2 and GP3-1 was determined by the same method as that used for GP4-4: Thr-Ser-Glu. GP3-1 was distinguished by a molecular weight approximately twice that of GP5-2 as judged by their behavior on gel filtration chromatography. This result, together with the analytical data based on SIMS and ^{1}H NMR measurements, establishes that the structures labeled C and D in Figure 3 can be assigned to GP5-2 and GP3-1, respectively.

SIMS Spectra of GP6-1, GP5-1, and GP5-2. In Figure 4 the positive-ion SIMS spectrum of the underivatized glycopeptide, GP6-1, is shown. The spectrum appears to consist of the molecular ion (plus pseudo molecular ions), its fragments, and cluster ions of glycerol [$(nG + H)^+$]. In the high-mass region, peaks at m/z 748, $(M + Na)^+$, and m/z 770, $(M + 2Na - H)^+$, are seen in addition to the peak, $(M + H)^+$, at m/z 726. We assume that the former two peaks are the result of multiple sodium addition, which is often observed in the positive-ion molecular SIMS spectra (Kambara & Hishida, 1981; Busch et al., 1982). These ions are compatible with the calculated molecular weight (M_r 725) of GP6-1 on the basis of composition. Fragment ions are formed by the rupture of the bonds along the peptide backbone producing diagnostic sequence ions at m/z 528, 506, 484, 267, 245, 229, and 132. The mass of the prominent ions at m/z 506 and 528 corresponds to the mass of the amino-terminal Thr residue to which the short-core oligosaccharide chain is attached. Sites of bond cleavages leading to the production of the primary sequence ions are depicted in Figure 4. Thus, the complete sequence of GP6-1 could be deduced directly from the SIMS spectrum with a series of characteristic fragment ions.

The positive-ion molecular SIMS spectrum of GP5-1 (Figure 5A) again showed prominent pseudo molecular ions responsible for $(M + H)^+$, $(M + Na)^+$, and $(M + 2Na - H)^+$ at m/z 1237, 1259, and 1281, respectively. These ions are in good agreement with the molecular weight for GP5-1 on the basis of the established structure (Figure 3) for this molecule. In this instance, the m/z 100–400 mass region of the spectrum gave no additional useful information on the primary structure. The presence of a Fuc-containing oligosaccharide unit is supported by the appearance of the peaks at m/z 674 and 512, characteristic of the sequence ions (Fuc-*O*-GalNAc-*O*-Gal-*O*-Gal)$^+$ and (Fuc-*O*-GalNAc-*O*-Gal)$^+$, respectively (Shimamura et al., 1983, 1984a).

As a third example, we shall give the SIMS spectrum of GP5-2 in Figure 5B. Positive molecular and pseudo molecular

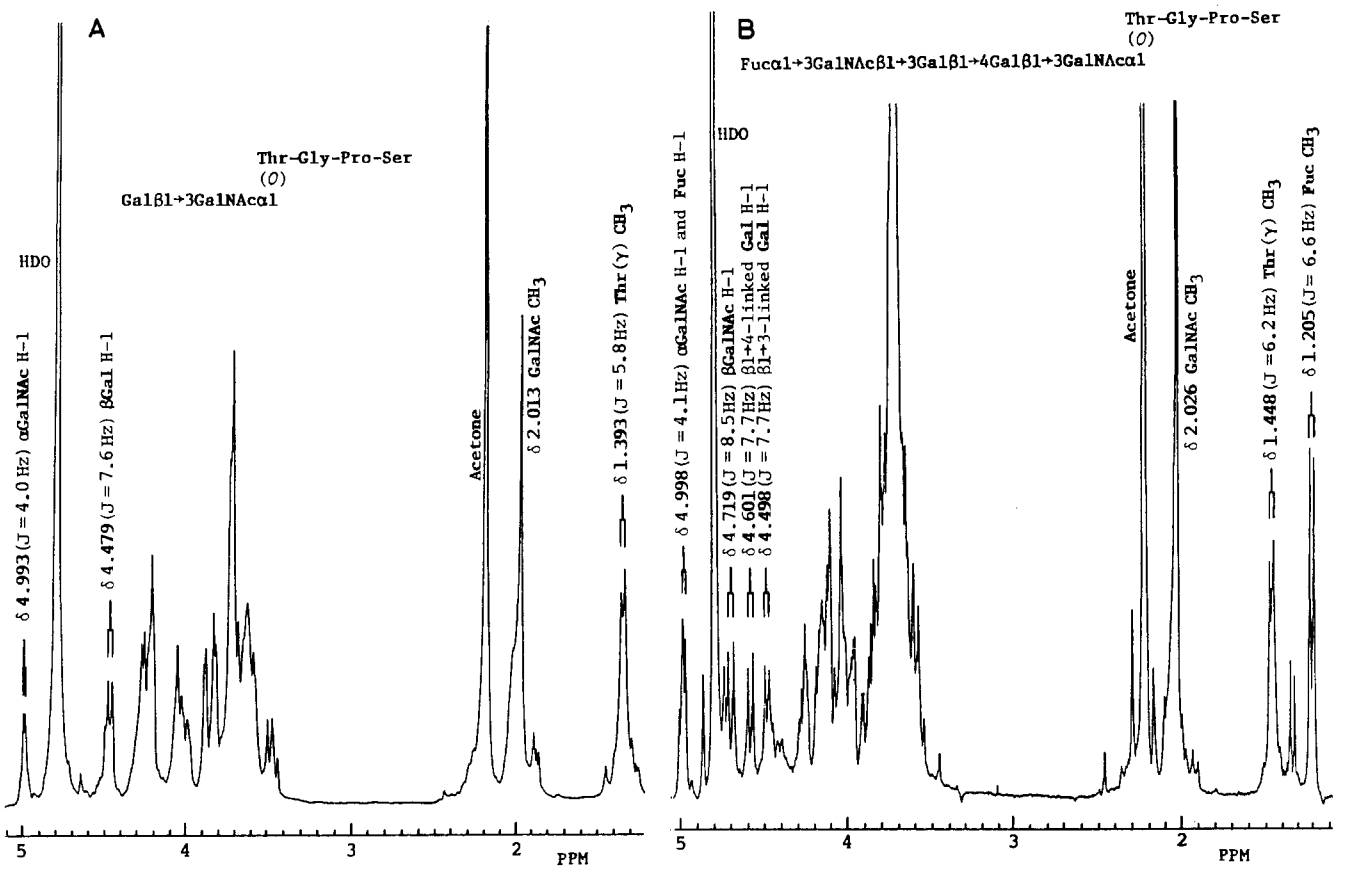


FIGURE 6: The 270-MHz ^1H NMR spectra of (A) GP6-1 and (B) GP5-1 in D_2O at 23 °C.

ions are apparent at m/z 1066, $(M + H)^+$, 1088, $(M + Na)^+$, and 1100, $(M + 2Na - H)^+$, confirming that the major component of GP5-2 is structure C in Figure 3.

The 270-MHz ^1H NMR Spectra of Glycotri- and Glycotetrapeptides. The 270-MHz ^1H NMR spectra of GP6-1 and GP5-1, recorded in D_2O at 23 °C and pD 7, are given in Figure 6. Inspection of the spectrum of GP6-1 allows two anomeric protons of the carbohydrate moiety to be readily assigned to individual monosaccharide residues due to their intensity and unique presence in characteristic regions: δ 4.993 ppm ($J_{1,2}$ = 4.0 Hz), αGalNAc H-1; δ 4.479 ppm ($J_{1,2}$ = 7.6 Hz), βGal H-1. The latter signal can be compared with the corresponding values for the βGal residue in Galβ1→3GalNAcol [δ 4.475 ppm, $J_{1,2}$ = 7.3 Hz at 25 °C (van Halbeek et al., 1980)]. On close comparison of the spectra of GP5-1 and an oligosaccharide alditol, Fucα1→3GalNAcβ1→3Galβ1→4Galβ1→3GalNAcol (Shimamura et al., 1983), an overlapped pair of doublets at δ ~4.998 ppm ($J_{1,2}$ = 4.1 Hz) can be assigned to anomeric protons of the terminal Fuc and proximal GalNAc residues. Three other well-resolved proton signals of equal intensity seen at δ 4.719 ppm ($J_{1,2}$ = 8.5 Hz), δ 4.601 ppm ($J_{1,2}$ = 7.7 Hz), and δ 4.498 ppm ($J_{1,2}$ = 7.7 Hz) can be readily assigned to the anomeric protons of the penultimate βGalNAc, β1→4-linked Gal, and β1→3-linked Gal residues, respectively, thereby the nature of the glycosidic bond between the monosaccharides being confirmed unequivocally.

The 270-MHz NMR spectra of GP5-2 and GP3-1 are shown in Figure 7. Inspection of the anomeric region of GP5-2 (Figure 7A) reveals two pairs of relatively well-resolved signals of equal intensity. A pair of signals with $J_{1,2}$ = 3.7 Hz appearing at δ 4.975 ppm and δ 4.924 ppm represent the two α-linked GalNAc residues attached to the peptide via Thr and Ser residues. For GalNAc H-1 in the α-*O*-Thr linkage, the expected chemical shift is δ ~4.99 ppm as was for GP6-1 and GP5-1, so that we can safely assign the downfield resonance (δ 4.975 ppm) to the αGalNAc-Thr anomeric proton and the upfield resonance (δ 4.924 ppm) to the αGalNAc-Ser anomeric proton. Another pair of signals with $J_{1,2}$ = 7.7 Hz stem from the two Gal residues: δ 4.472 ppm, the terminal Gal H-1 of the short-core oligosaccharide attached to Thr residue, and δ 4.516 ppm, the corresponding anomeric proton of another short core linked to Ser residue. Thus, the ^1H NMR spectrum of GP5-2 (Figure 7A) fully supports structure C in Figure 3 and the anomeric nature of the four glycosides. Similarly, with the structure of GP3-1 established (Figure 3), all of the anomeric proton resonances can be assigned to individual monosaccharides on the basis of information newly obtained for GP6-1, GP5-1, and GP5-2 and of previously reported chemical shift data of Fuc-containing pentasaccharide alditol (Shimamura et al., 1983). The chemical shifts and coupling constants of the selected protons in six glycopeptides are summarized in Table III.

The 270-MHz spectrum of GP4-4 is given in Figure 8. The three anomeric proton signals at δ 4.997 ppm ($J_{1,2}$ = 4.3 Hz), δ 4.970 ppm ($J_{1,2}$ = 3.9 Hz), and δ 4.924 ppm ($J_{1,2}$ = 3.7 Hz) must stem from the glycosidic residues of which anomeric protons are in an equatorial position and coupled with the neighboring axial H-2; it therefore follows that the lowest field resonance represents the αFuc H-1 and the latter two signals are assigned to the Thr-linked αGalNAc and Ser-linked αGalNAc residues, respectively, solely on the basis of a comparison of their chemical shifts with those of GP3-1 and GP5-2. Since GP4-4 contains two carbohydrate units, one short core and one Fuc-containing core, which are located closely spaced within a molecule, the prominent signals appearing at δ 4.724 ppm ($J_{1,2}$ = 8.2 Hz) and at δ 4.603 ppm ($J_{1,2}$ = 7.8 Hz) are readily assigned to the anomeric protons of the penultimate GalNAc residue and the central β1→4-

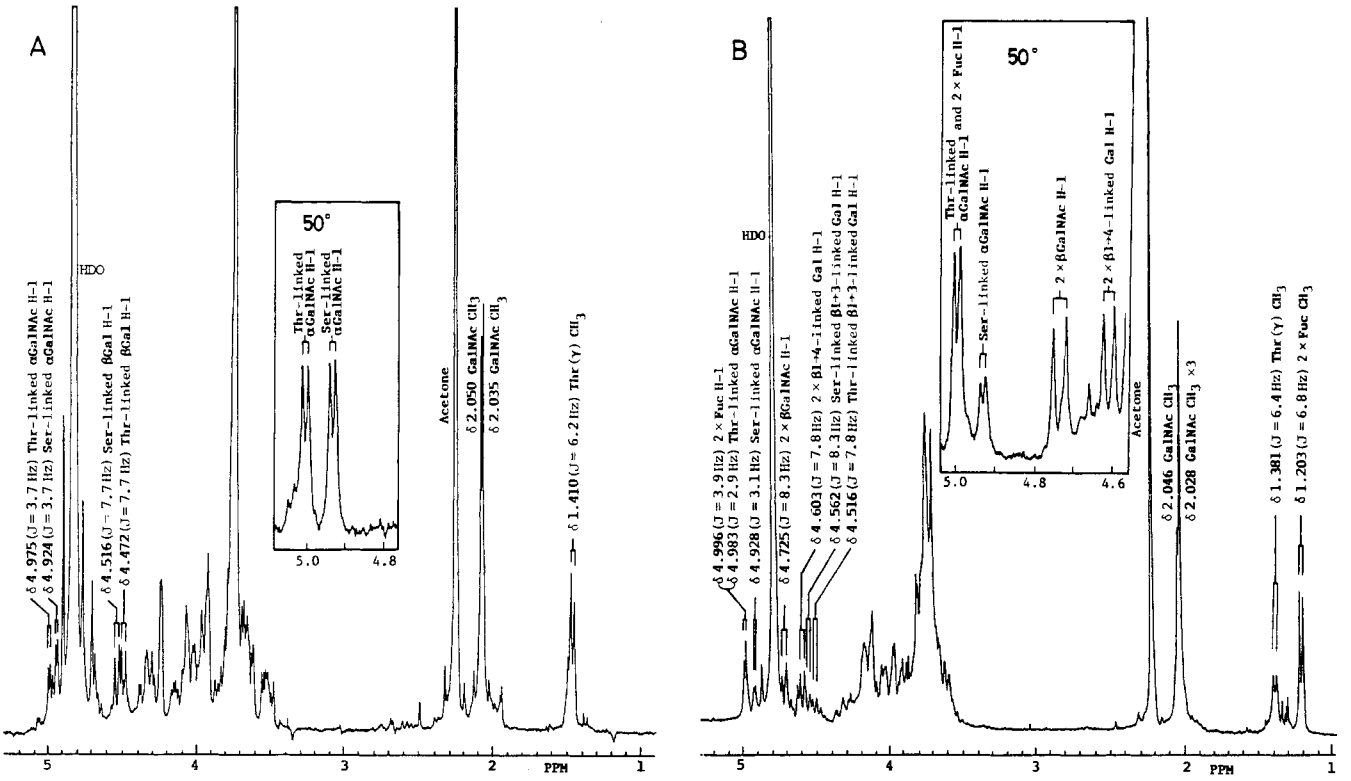


FIGURE 7: The 270-MHz ^1H NMR spectra of (A) GP5-2 and (B) GP3-1 in D_2O at 23 °C. The anomeric region of the spectra at 50 °C is also shown.

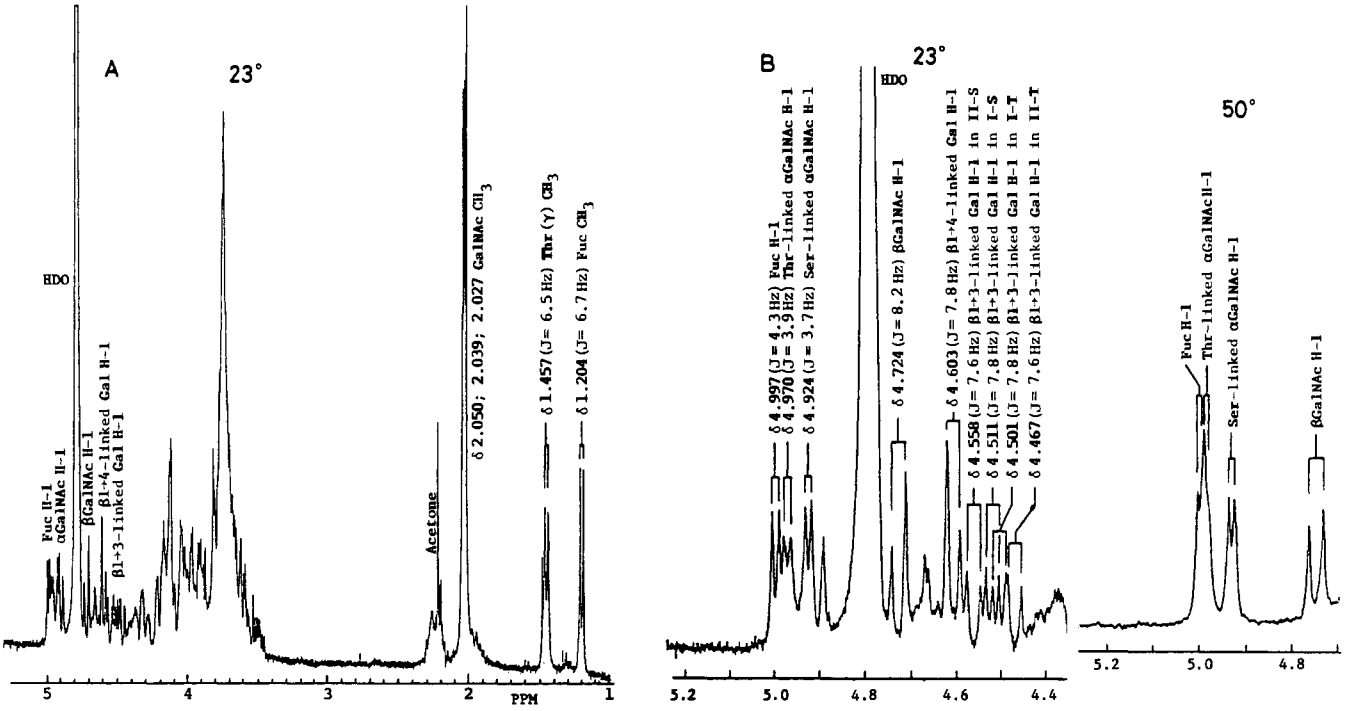


FIGURE 8: (A) ^1H NMR spectrum at 270 MHz of GP4-4 at 23 °C in D_2O; (B) expanded spectra comparing the anomeric region at 23 and 50 °C. The residue notation in structures I and II is as follows:

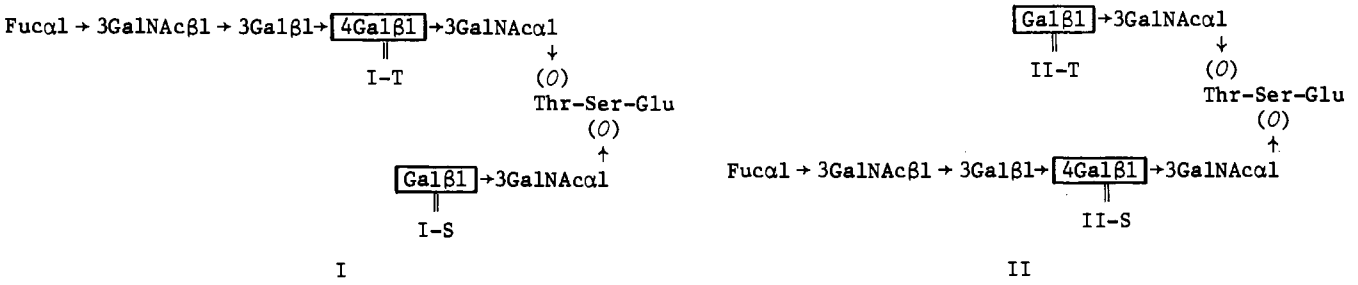


Table III: Proton (Anomeric, Fuc Methyl, *N*-Acetyl, and Thr γ-Methyl) Chemical Shifts of Glycopeptides[a]

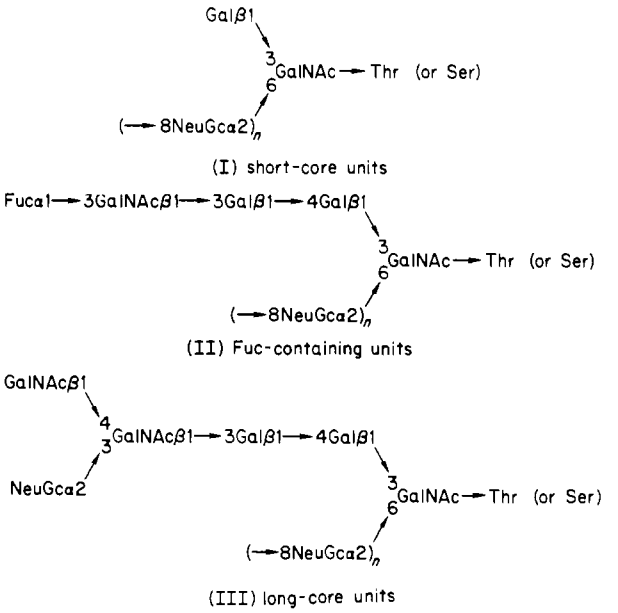
proton	GP6-1	GP5-1	GP5-2	GP4-4	GP3-1
H-1 of					
terminal αFuc		4.998 (4.1)		4.997 (4.3)	4.996 (3.9)
αGalNAc in Thr-linked chain	4.993 (4.0)	4.998 (4.1) [5.001 (4.2)]	4.975 (3.7) [5.003 (3.7)]	4.970 (3.9)[4.987 (3.9)]	4.983 (2.9) [5.000 (3.9)]
αGalNAc in Ser-linked chain			4.924 (3.7) [4.935 (3.5)]	4.924 (3.7) [4.931 (3.4)]	4.928 (3.1) [4.935 (3.7)]
penultimate βGalNAc		4.719 (8.5) [4.746 (8.3)]		4.724 (8.2) [4.748 (8.3)]	4.725 (8.3) [4.750 (8.3)]
β1→3-linked Gal in Thr-linked chain	4.479 (7.6)	4.498 (7.7)	4.472 (7.7)	4.501 (7.8) (I-T)[b] 4.467 (7.6) (II-T)[b]	4.516 (7.8)
β1→3-linked Gal in Ser-linked chain			4.516 (7.7)	4.558 (7.6) (II-S)[b] 4.511 (7.8) (I-S)[b]	4.562 (8.3)
β1→4-linked Gal		4.601 (7.7)		4.603 (7.8)	4.603 (7.8)
CH_3 of					
terminal αFuc		1.205 (6.6)		1.204 (6.7)	1.203 (6.8)
GalNAc	2.013	2.026 × 2	2.035, 2.050	2.027, 2.039, 2.050	2.028 × 3, 2.046
Thr γ-methyl	1.393 (5.8)	1.448 (6.2)	1.410 (6.2)	1.457 (6.5)	1.381 (6.4)

[a] Chemical shifts are measured in D_2O at 23 °C and are expressed in ppm from DSS; values in parentheses are $J_{1,2}$ in hertz; those in brackets are the corresponding values at 50 °C. [b] The notation designating the position of the residue is shown in the structures in Figure 8.

linked Gal residue on the Fuc-containing sugar chain. These residues are distant from the peptide terminals and oligosaccharide attachment sites, so that their H-1 chemical shifts are not affected by the possible difference in the attachment of fucopentasaccharide to the peptide. A group of signals between δ 4.45 ppm and δ 4.57 ppm with a combined intensity comparable to that of two proton resonances can be assigned to the anomeric protons of the four different β1→3-linked Gal residues: First, we can assign the upfield resonance at δ 4.467 ppm ($J_{1,2}$ = 7.6 Hz) to the β1→3-linked Gal H-1 of the Thr-linked short core and the downfield resonance (δ 4.558 ppm, $J_{1,2}$ = 7.6 Hz) to the β1→3-linked Gal H-1 of the Ser-linked Fuc-containing core. Second, assignments for the other two signals have been made on the basis of a comparison of their chemical shifts with those estimated empirically by assuming the substitution effects on the chemical shifts to be additive [cf. Vliegenthart et al. (1983)]: δ for the β1→3-linked Gal H-1 of the Thr-linked Fuc-containing core (I-T) = δ for II-T + (δ for GP5-1 − δ for GP6-1) = 4.49 ppm, and δ for the β1→3-linked Gal H-1 of the Ser-linked short core (I-S) = δ for II-S − [δ for GP3-1(Ser) − δ for GP5-2(Ser)] = 4.51 ppm. Experimentally, two resonances of equal intensity were observed at δ 4.501 ppm ($J_{1,2}$ = 7.8 Hz) and δ 4.511 ppm ($J_{1,2}$ = 7.8 Hz). We can thus assign the former resonance to the β1→3-linked Gal H-1 of the Thr-linked Fuc-containing core (I-T) and the latter resonance to the β1→3-linked Gal H-1 of the Ser-linked short core (I-S). It should be noted that on the basis of the relative intensity of the observed resonances (I-T vs. II-T or I-S vs. II-S) being found in approximately 1:1 ratio the glycopeptide GP4-4 must be a 1:1 mixture of structures E and F depicted in Figure 3. This was substantiated by the fact that GP4-4, when subjected to BH_4^-/OH^- treatment followed by thin-layer chromatography, showed two spots corresponding to Galβ1→3GalNAcol and Fucα1→3GalNAcβ1→3Galβ1→4Galβ1→3GalNAcol, both of which being derived from the β-elimination-sensitive carbohydrate moiety located at the internal Ser residue.

Discussion

Since salmon egg PSGP has a weight-average molecular weight of ~200 000 and since it contains about 85% (w/w) of the total carbohydrate and ~60% (w/w) of sialic acid (NeuGc), the fact that PSGP runs as a broad band on SDS–polyacrylamide gel electrophoresis is presumably due to heterogeneity in the carbohydrate units. Previous studies from this laboratory (Shimamura et al., 1983, 1984a) have demonstrated that the salmon egg PSGP contains three types of homologous carbohydrate units:



The relative proportions of the three types of core oligosaccharides present in salmon egg PSGP were found to be (short-core unit) 8.3, (Fuc-containing unit) 7.5, and (long-core unit) 1.0 (Shimamura et al., 1984a). Although the actual role of PSGP is not yet understood, its polyanionic nature suggests a possible role in intracellular Ca^{2+} ion and/or Na^+/H^+ ion homeostasis at fertilization and during subsequent development. In view of this, the unique oligo- and/or polysialylglycan side chains are believed to be where the function resides, but at the same time any functional model must take into account the structure of the entire molecule of PSGP.

As an extension of the studies on carbohydrate prosthetic groups, we felt it necessary to determine the amino acid sequence of the core protein. First, we have been particularly interested in isolating glycopeptides for the determination of the amino acid sequences surrounding the glycosylated amino acid residues. Purification and analysis of major glycopeptides derived from desialylated PSGP after a number of fractionation steps have shown that the number of glycopeptides produced is extremely limited considering the large number of sites (~60) of attachment of the carbohydrate units to the core protein in a single molecule of PSGP. The oligopeptide structures of the glycopeptides belong to only two distinct groups: Thr*-Ser*-Glu and Thr*-Gly-Pro-Ser. Thus, several

of the major glycopeptides share the common peptide sequences Thr-Ser-Glu and Thr-Gly-Pro-Ser, though microheterogeneity is apparent in the carbohydrate moieties. The two abundant types of carbohydrate units, i.e., short-core and Fuc-containing units, represent about 94% of the total carbohydrate units of salmon egg PSGP. On the basis of the yield and composition of each of the glycotri- and glycotetrapeptides given in Table I, it is apparent that the two different classes of carbohydrate units are rather randomly distributed at the points of their attachment along the peptide chain.

The weight of glycosylated tri- and tetrapeptides released from desialylated PSGP after exhaustive hydrolysis by Pronase was at least 64%, which would account for two-thirds of the carbohydrate binding sites in PSGP. This shows that Pronase attack on desialylated PSGP would result in the removal of internal Ala and Asp residues, which are present in the Pronase-susceptible region of the molecule. The present results also suggest a high frequency of occurrence of -Thr*-Ser*-Glu- and -Thr*-Gly-Pro-Ser- sequences along the PSGP chain. The overall structural organization could consist of repetition of these two glycopeptide units that are thought to be interspersed with rather short nonglycosylated peptides comprising Ala and Asp residues.

It should be noted that only half of the total Ser residues in PSGP are glycosylated while all the Thr residues are glycosylated. The Ser residues in every -Thr-Gly-Pro-Ser- sequence, for unknown reason, have specifically been prevented from glycosylation.

Previous studies using α-*N*-acetylgalactosaminidase have demonstrated that the glycosidic linkages of GalNAc to Ser or Thr residues in the ovine submaxillary mucin (Buddecke et al., 1969), hog stomach blood group A substance (Weissman & Hinrichsen, 1969), and fetuin (Spiro & Bhoyroo, 1974) are all α in anomeric configuration. More recently, the anomeric form of the peptide-linked GalNAc residues in the blood group M and N active glycopeptides from human glycophorin A (Prohaska et al., 1981) and ovine submaxillary mucin (Gerken & Dearborn, 1984) was studied by ^{13}C NMR spectroscopy, and that in the B chain of human plasma α_2HS glycoprotein (Gejyo et al., 1983) was deduced to be α by ^1H NMR studies. Until now there was no direct evidence for the stereochemistry at the anomeric carbon of the Ser- and Thr-linked GalNAc residues of PSGP. The anomeric region of the ^1H NMR spectrum of, say, GP5-1 (Figure 5B) is nearly identical with that of Fucα1→3GalNAcβ1→3Galβ1→4Galβ1→3GalNAcol, the only difference being the presence of the signal at δ 4.998 ppm ($J_{1,2}$ = 4.1 Hz) in the former glycopeptide and the absence of this signal for the latter oligosaccharide alditol. This immediately led us to assign the signal at δ 4.998 ppm to the H-1 of the proximal GalNAc residue in GP5-1. The coupling constant $J_{1,2}$ = 4.1 Hz corresponds to an equatorial–axial coupling between H-1 and H-2; i.e., H-1 is in an equatorial position; it therefore follows that GalNAc unit is again joined by α-glycosidic linkage to Thr residue. Similarly, all the proximal GalNAc H-1 resonances in other glycopeptides studied are observed as doublets corresponding to the α-glycosidic configuration.

It is interesting to note that the two different types of the peptide-linked GalNAc residues, i.e., GalNAcα1→Ser and GalNAcα1→Thr, can apparently be distinguished on the basis of their H-1 chemical shift data: the GalNAc H-1 of the former type always resonates at somewhat higher field than that of the latter type (Table III). However, these results probably reflect either the environmental difference in the proximal GalNAc residues at the N-terminal Thr and the internal Ser residues or the intrinsic difference in the nature of the glycosidic bond between GalNAcα1→Ser and GalNAcα1→Thr. Inspection of the spectrum at 50 °C of GP2-5 (not shown),[2] in which the carbohydrate-attached Thr residue is no longer at the N-terminal position, reveals likewise the Thr-linked GalNAc H-1 resonance at δ 4.973 ppm and the Ser-linked GalNAc H-1 signal at δ 4.926 ppm. Furthermore, on the basis of the ^1H NMR data for the structurally related model glycopeptides reported in the literature, perBzlGlcα1→Ser(*N*-2,4-dinitrophenyl) methyl ester (δ 4.87 ppm) vs. perBzlGlcα1→Thr(*N*-2,4-dinitrophenyl) methyl ester (δ 5.05 ppm) (Vercellotti et al., 1970) and Fucα1→Ser (δ 4.88 ppm) vs. Fucα1→Thr (δ 5.05 ppm) (Matsue & Majima, 1983), we can safely assign the upfield resonance to the GalNAcα1→Ser anomeric proton and the downfield resonance to the GalNAcα1→Thr anomeric proton. Thus, the chemical shifts δ 4.97–5.01 ppm and δ 4.92–4.94 ppm can be used as diagnostic values for determination of the linkage amino acid in the mucin-type glycopeptides.

The methyl protons of the glycosylated Thr residues resonate at δ 1.38–1.46 ppm (J = 5.8–6.3 Hz). The methyl protons of the free Thr have been assigned to the resonances at about δ 1.32 ppm for its zwitterionic form and near δ 1.36 ppm for its cationic form (at pH 0.5). This, taken together with other data, shows that the attachment of a carbohydrate chain to a Thr residue through the GalNAc residue gives rise to a downfield shift of the Thr γ-methyl proton signal.

Finally, in the present study, amino acid sequences around the carbohydrate linkage sites have been firmly established for salmon egg PSGP. It should be emphasized that the core protein in the PSGP molecule is considered to be made up of repeating sequences of the two distinct types of glycopeptides, Thr*-Ser*-Glu and Thr*-Gly-Pro-Ser. In this connection, the complete sequence of the amino acids in trout egg PSGP is under current study, and possible implications of the occurrence of such repetitive sequences for the functions of this novel class of glycoproteins will be discussed elsewhere.

ACKNOWLEDGMENTS

We thank Professor T. Miyazawa, Dr. T. Endo, and T. Watanabe of the University of Tokyo for their help in obtaining the ^1H NMR spectra. We are also indebted to Drs. H. Kambara and S. Seki of Central Research Laboratory, Hitachi, Ltd., who generously conducted the measurements of SIMS spectra.

REFERENCES

Bennett, J. C. (1967) *Methods Enzymol. 11*, 211–231.

Buddecke, E., Schauer, R., Werries, E., & Gottschalk, A. (1969) *Biochem. Biophys. Res. Commun. 34*, 517–521.

Busch, K. L., Unger, S. E., Vincze, A., Cooks, R. G., & Keough, T. (1982) *J. Am. Chem. Soc. 104*, 1507–1511.

Derevitskaya, V. A., Vafina, M. G., & Kochetkov, N. K. (1967) *Carbohydr. Res. 3*, 377–388.

Dubois, M., Gilles, K. A., Hamilton, J. K., Rebers, P. A., & Smith, F. (1956) *Anal. Chem. 135*, 243–250.

Gejyo, F., Chang, J.-L., Bürgi, W., Schmid, K., Offner, G. D., Troxler, R. F., van Halbeek, H., Dorland, L., Gerwig,

[2] Anion-exchange borate chromatography of GP2 yielded several phenol–sulfuric acid positive peaks. These minor components were also found to contain NeuGc residues. Of them, the glycopeptide named GP2-5 appeared to be homogeneous enough, as judged from the amino acid analysis, to be sequenced, and the structure deduced is Asp-Asp-Ala-Thr*-Ser*-Glu. Although a full analysis of the structure for GP2-5 could not be made because of an insufficient amount, a carbohydrate unit attached to the Thr residue might sterically block access to the particular local peptide sequence -Asp-Asp-*Ala-Thr**-Ser*-Glu- by Pronase.

G. J., & Vliegenthart, J. F. G. (1983) *J. Biol. Chem. 258*, 4966–4971.
Gerken, T. A., & Dearborn, D. G. (1984) *Biochemistry 23*, 1485–1497.
Gilkey, J. C. (1981) *Am. Zool. 21*, 359–375.
Gray, W. R. (1972a) *Methods Enzymol. 25*, 121–138.
Gray, W. R. (1972b) *Methods Enzymol. 25*, 333–344.
Hayashi, R., Bai, Y., & Hata, T. (1975) *J. Biochem. (Tokyo) 77*, 69–79.
Inoue, S., & Iwasaki, M. (1978) *Biochem. Biophys. Res. Commun. 83*, 1018–1023.
Inoue, S., & Matsumura, G. (1979) *Carbohydr. Res. 74*, 361–368.
Inoue, S., & Iwasaki, M. (1980) *Biochem. Biophys. Res. Commun. 93*, 162–165.
Inoue, S., Iwasaki, M., & Matsumura, G. (1981) *Biochem. Biophys. Res. Commun. 102*, 1295–1301.
Iwasaki, M., Inoue, S., Kitajima, K., Nomoto, H., & Inoue, Y. (1984a) *Biochemistry 23*, 305–310.
Iwasaki, M., Nomoto, H., Kitajima, K., Inoue, S., & Inoue, Y. (1984b) *Biochem. Int. 8*, 573–579.
Kambara, H., & Hishida, S. (1981) *Anal. Chem. 53*, 2340–2344.
Kitajima, K., Nomoto, H., Inoue, Y., Iwasaki, M., & Inoue, S. (1984) *Biochemistry 23*, 310–316.
Kozulić, B., Ries, B., & Mildner, P. (1979) *Anal. Biochem. 94*, 36–39.
Laemmli, U. K. (1970) *Nature (London) 227*, 680–685.
Matsue, H., & Majima, M. (1983) *J. Biochem. (Tokyo) 94*, 1085–1094.
Moore, S., & Stein, W. H. (1954) *J. Biol. Chem. 211*, 907–913.
Morse, D., & Horecker, B. L. (1966) *Anal. Biochem. 14*, 429–433.
Nomoto, H., & Inoue, Y. (1983) *Eur. J. Biochem. 135*, 243–250.
Nomoto, H., Iwasaki, M., Endo, T., Inoue, S., Inoue, Y., & Matsumura, G. (1982) *Arch. Biochem. Biophys. 218*, 335–341.
Pigman, W., Moschera, J., Downs, F., & Wakabayashi, T. (1974) in *Actes du Colloque International No. 221 du Centre National de la Recherche Scientifique sur les Glycoconjugues* (Montreuil, J., Ed.) Vol. 1, pp 231–247, CNRS, Paris.
Prohaska, R., Koerner, A. W., Jr., Armitage, I. M., & Furthmayr, H. (1981) *J. Biol. Chem. 256*, 5781–5791.
Reinhold, V. N. (1972) *Methods Enzymol. 25*, 244–249.
Shimamura, M., Endo, T., Inoue, Y., & Inoue, S. (1983) *Biochemistry 22*, 959–963.
Shimamura, M., Endo, T., Inoue, Y., Inoue, S., & Kambara, H. (1984a) *Biochemistry 23*, 317–322.
Shimamura, M., Inoue, Y., & Inoue, S. (1984b) *Arch. Biochem. Biophys. 232*, 699–706.
Spiro, R. G. (1966) *Methods Enzymol. 8*, 26–52.
Spiro, R. G., & Bhoyroo, V. D. (1974) *J. Biol. Chem. 249*, 5704–5717.
Tomita, M., Furthmayr, H., & Marchesi, V. (1978) *Biochemistry 17*, 4756–4770.
Van Halbeek, H., Dorland, L., Vliegenthart, J. F. G., Fiat, A.-M., & Jolles, P. (1980) *Biochim. Biophys. Acta 623*, 295–300.
Vercellotti, J. R., Nienaber, N., & Chang, C. J. (1970) *Carbohydr. Res. 13*, 63–74.
Vliegenthart, J. F. G., Dorland, L., & Van Halbeek, H. (1983) *Adv. Carbohydr. Chem. Biochem. 41*, 209–374.
Weissman, B., & Hinrichsen, D. F. (1969) *Biochemistry 8*, 2034–2043.
Yphantis, D. A. (1960) *Ann. N.Y. Acad. Sci. 88*, 586–601.

Isolation and Characterization of Angiogenin, an Angiogenic Protein from Human Carcinoma Cells†

James W. Fett, Daniel J. Strydom, Roy R. Lobb, Edward M. Alderman, J. Lemuel Bethune, James F. Riordan, and Bert L. Vallee*

Center for Biochemical and Biophysical Sciences and Medicine and Departments of Biological Chemistry, Pathology, and Radiology, Harvard Medical School, and Brigham and Women's Hospital, Boston, Massachusetts 02115

Received May 9, 1985

ABSTRACT: The first human tumor derived protein with in vivo angiogenic activity to be obtained in pure form has been isolated from serum-free supernatants of an established human adenocarcinoma cell line (HT-29) and named angiogenin. It was purified by cation-exchange and reversed-phase high-performance liquid chromatography; the yield was ~0.5 μg/L of medium. Biological activity of angiogenin was monitored throughout purification by using the chick embryo chorioallantoic membrane assay. Statistical evaluation demonstrates that it displays activity in this system with as little as 35 fmol per egg. Moreover, only 3.5 pmol is required to induce extensive blood vessel growth in the rabbit cornea. The amino acid composition of this basic (isoelectric point >9.5), single-chain protein of molecular weight ~14 400 has been determined. The amino terminus is blocked, and the carboxyl-terminal residue is proline.

Angiogenesis, the induction of the formation of blood vessels, is critical to the development, progression, and metastasis of animal and human tumors (Goldman, 1907; Ide et al., 1939; Algire & Chalkey, 1945; Ehrmann & Knoth, 1968; Greenblatt & Shubik, 1968; Folkman & Cotran, 1976). The molecular messengers that arise from tumor cells to mediate this process have long been sought (Folkman et al., 1971; Tuan et al., 1973; Phillips & Kumar, 1979; Weiss et al., 1979; McAuslan, 1980; Fenselau et al., 1981; Kumar et al., 1983; McAuslan et al.,

† This work was supported by funds from the Monsanto Co. under agreements with Harvard University.

* Address correspondence to this author at the Center for Biochemical and Biophysical Sciences and Medicine, Harvard Medical School.

0006-2960/85/0424-5480$01.50/0