Carbohydrate Research, 120 (1983) 85–94
Elsevier Science Publishers B.V., Amsterdam – Printed in The Netherlands

SPECIFIC METHOD FOR THE FRAGMENTATION OF THE POLYPEPTIDE CHAIN OF GLYCOPROTEINS. DISTRIBUTION OF CARBOHYDRATE CHAINS ON THE PEPTIDE CORE OF BLOOD-GROUP-SPECIFIC GLYCOPROTEIN*

VARVARA A. DEREVITSKAYA, L. M. LIKHOSHERSTOV, M. D. MARTYNOVA, AND NIKOLAI K. KOCHETKOV

N.D. Zelinsky Institute of Organic Chemistry, Academy of Sciences of the U.S.S.R., Moscow B-334 (U.S.S.R.)

(Received December 13th, 1982; accepted for publication, May 2nd, 1983)

ABSTRACT

A method for specific fragmentation of the polypeptide backbone of glycoproteins at the glycosylated serine and threonine residues has been developed. The fragmentation includes β-elimination of the carbohydrate chains, followed by bromination of the resulting enamine groups, and cleavage of the brominated amino acid residues by alkaline sodium borohydride. The method was employed for fragmentation of the peptide core of pig blood-group substance H. Essentially all the serine and threonine residues were shown to be *O*-glycosylated, and rather frequently either adjacent or separated by a single amino acid (mainly alanine). When they were separated by two or three amino acid residues, proline was preponderant.

INTRODUCTION

Much attention has been devoted to the investigation of the structure of blood-group substances, mainly to the carbohydrate chains of these complicated glycoproteins[1,2]. Still little is known of the peptide-core structure and carbohydrate-chain distribution on the peptide backbone[3,4]. This information would be of value for our understanding of the spatial structure of these important glycoproteins.

Dehydroalanine and 2-aminobutenoic acid residues that are formed by the β-elimination of carbohydrate chains under mild alkaline conditions contain enaminic groups which may be suitable points for fragmentation of the polypeptide chain. Sokolovsky and Patchornik[5] have reported fragmentation of several synthetic dehydroalanyl peptides by alkaline treatment after bromination or mild acid hydrolysis. The latter procedure was found to be unsuitable for peptides that con-

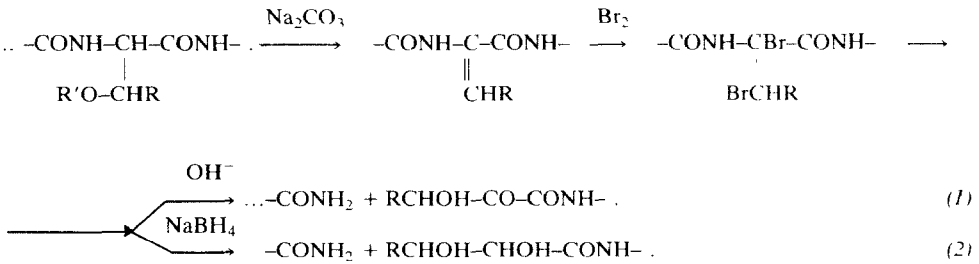
*Dedicated to Professor Elvin A. Kabat.

0008-6215/83/$ 03.00

tain 2-aminobutenoic acid residues[6]. In the first attempt to cleave the blood-group-glycoprotein peptide chains by the method of Sokolovsky and Patchornik[5], extensive depolymerisation of the glycoprotein was observed[7]. The same method was used later for the isolation of minor, *N*-linked glycopeptides of blood-group glycoproteins[8].

In the present paper, we describe a new method for the specific cleavage of the peptide chain of *O*-glycoproteins. This method was used for the cleavage of blood-group H glycoprotein, giving stable fragments for fractionation and analysis. Some conclusions on the distribution of the carbohydrate chains on the polypeptide core were drawn from the results of the fragment analysis.

RESULTS AND DISCUSSION



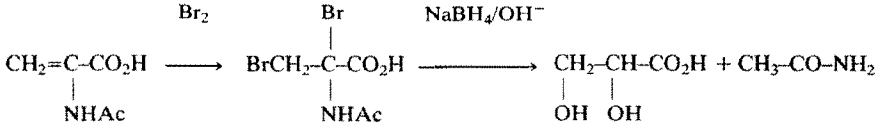
R = H or Me
R' = carbohydrate chain

Scheme 1

Scheme 1 describes the specific cleavage of the peptide chains in glycoproteins according to the procedure suggested by Sokolovsky and Patchornik[5] (*1*), and by the method presented herein (*2*). The reactions involve alkaline elimination of the carbohydrate chains from the peptide core, bromination of the resulting enaminic acids, and alkaline cleavage of the peptide backbone. Thus, Procedure *1* results in the formation of amides at the *C*-terminus, and 3-hydroxy-2-oxopropanoic acid (from serine) or 3-hydroxy-2-oxobutanoic acid (from threonine) residues at the *N*-terminus of the fragments obtained. Attempts to apply Procedure *1* to the cleavage of blood-group substance H showed some difficulties in the analysis of *N*-terminal groups, probably because of the lability of these groups.

In order to obtain stable products, the peptide chain was cleaved with alkaline sodium borohydride after bromination. This resulted in the conversion of the 3-hydroxy-2-oxo acid into 2,3-dihydroxy acid residues (Procedure *2*, Scheme 1). The conditions of cleavage were selected by varying the concentration of bromine and sodium borohydride, time, and pH, and using *N*-acetyldehydroalanine as a model compound (Scheme 2). The yield of glyceric acid was determined by estimation of the formaldehyde liberated after periodate oxidation of the reaction products. Under optimal conditions (excess of bromine; 0.4M sodium

borohydride, 4 h, pH 11), the yield of glyceric acid was ~90%. Glyceric acid and acetamide were identified by paper electrophoresis and paper chromatography.



Scheme 2

Method *2* was applied to the cleaving of the blood-group H peptide core. Under these conditions, formation of stable peptides containing glyceric acid residues (from serine) or 2,3-dihydroxybutanoic acid residues (from threonine) at the *N*-terminus may be expected. These terminal residues may be conveniently determined by quantitative determination of formaldehyde and acetaldehyde released by periodate oxidation. To eliminate the carbohydrate chains, the glycoprotein was subjected twice to the treatment with 50mM sodium carbonate for 4 h at 70°. These conditions were chosen to minimize the cleavage of peptide bonds, as ~50% of the carbohydrate components were split off without significant degradation of the peptide chain, as shown by the observation that the preponderant part of the alkali-treated product containing amino acids residues was excluded from a Bio-Gel P-100 column. Moreover, practically no amino acid residue could be detected in the

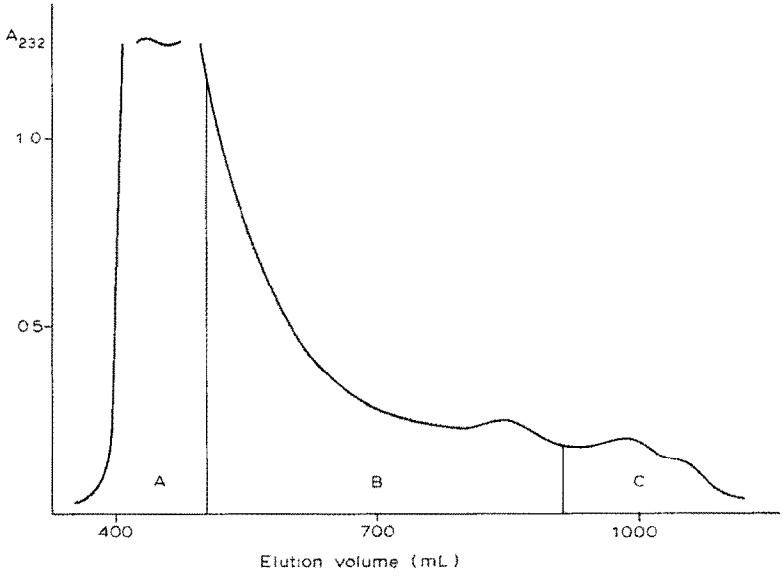


Fig. 1. Gel-filtration pattern of the cleavage products of the high-molecular-weight fraction on a Sephadex G-25 column (1.8 × 120 cm). The column was equilibrated and eluted with 0.1M acetic acid, and the absorbance at 232 nm determined.

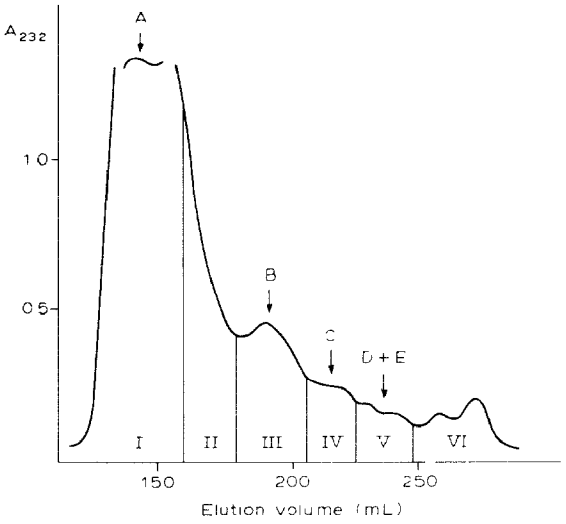


Fig. 2. Gel-filtration pattern of Fractions B plus C (from Fig. 1) on a column (1.4 × 230 cm) of Sephadex G-15 eluted with 0.1M acetic acid and monitored by absorbance at 232 nm. The arrows indicate the elution volumes of the reference compounds (molecular mass given in parentheses): (A) Blood-group glycoprotein (>300 000), (B) *N*-acetylhexaglycylamide (401), (C) tetraglycine (246), (D) diglycylamide (131), and (E) glycine (75).

TABLE I

AMINO ACID CONTENT OF HIGH-MOLECULAR-WEIGHT FRACTIONS, AND OF FRACTIONS ELUTED FROM SEPHADEX G-25 COLUMN (A) AND SEPHADEX G-15 COLUMN (I–VI)[a]

Amino acid	*Amount in fraction*							
	HMF	*A*	*I*	*II*	*III*	*IV*	*V*	*VI*
Aspartic acid	57.2	39.7	11.8	2.3	[b]	[c]	[c]	[c]
Threonine	334.0	247.0	69.0	11.4	2.1	[b]	[c]	[c]
Serine	128.0	85.6	22.9	3.2	[b]	[c]	[c]	[c]
Glutamic acid	111.4	65.9	23.5	6.0	3.2	[b]	2.4	[c]
Proline	507.6	323.1	112.5	20.5	14.8	5.3	3.1	[b]
Glycine	120.7	70.3	17.3	3.5	2.5	2.7	5.5	1.8
Alanine	204.6	108.4	35.7	8.1	8.3	5.3	8.3	2.1
Valine	137.0	77.3	28.6	5.2	4.0	1.5	2.7	[c]
Isoleucine	39.7	15.9	6.8	[b]	[c]	[c]	[c]	[c]
Leucine	88.7	44.9	13.0	3.0	3.2	1.5	2.0	[b]
Histidine	5.9	5.0	[b]	[c]	[c]	[c]	[c]	[c]
Lysine	13.0	15.1	2.2	[c]	[c]	[c]	[c]	[c]
Arginine	10.5	6.7	2.4	[c]	[c]	[c]	[c]	[c]
Total amino acids (Σ AA)	1758.3	1104.9	345.7	63.2	38.1	16.3	24.0	3.9

[a] μmol/g of high-molecular-weight fraction (abrev. HMF). [b] 1.0–1.5 μmol/g. [c] <1.0 μmol/g.

TABLE II

CONTENT OF GLYCERIC ACID, 2,3-DIHYDROXYBUTANOIC ACID, AND CARBOHYDRATE COMPONENTS OF THE HIGH-MOLECULAR-WEIGHT FRACTION (HMF), AND OF FRACTIONS ELUTED FROM SEPHADEX G-25 COLUMN (A) AND SEPHADEX G-15 COLUMN (I–VI)[a]

Component	*Amount in fraction*							
	HMF	*A*	*I*	*II*	*III*	*IV*	*V*	*VI*
Glyceric acid		34.8	20.1	4.8	6.0	4.8	16.1	3.9
2,3-Dihydroxybutanoic acid		62.5	44.6	11.1	10.6	11.7	33.0	11.8
Sum of dihydroxy acids		97.3	64.7	15.9	16.6	16.5	49.1	15.7
2-Amino-2-deoxy-D-glucose	904.4	699.5	168.0	16.0	*b*	*c*	*c*	*c*
2-Amino-2-deoxy-D-galactose	441.5	341.7	95.3	11.8	*b*	*c*	*c*	*c*
D-Galactose	593.2	432.4	107.2	7.9	*c*	*c*	*c*	*c*
L-Fucose	278.0	191.5	53.1	4.1	*c*	*c*	*c*	*c*
Sum of amino acids/ sum of dihydroxy acids		11.4	5.3	4.0	2.3	1.0	0.5	0.25

[a] μmol/g of high-molecular-weight fraction. [b] <1.0–1.5 μmol/g. [c] <1.0 μmol/g.

low-molecular-weight fraction. The high-molecular-weight fraction, which contained enamine groups, was treated as shown in Scheme 1 (*2*). The products were separated by chromatography on Sephadex G-25 (Fig. 1) to give Fraction A, eluted with the void volume of a column, and Fractions B and C. These last two fractions were combined (after desalting of Fraction C) and subjected to chromatography on Sephadex G-15 (Fractions I–VI, Fig. 2), and the content of amino acids (Table I), hexosamines, neutral monosaccharides, and glyceric acid and 2,3-dihydroxybutanoic acid residues (Table II) was determined. The monosaccharide content was found to decrease sharply beginning from Fraction II and it was essentially nil in Fractions III–VI. This result suggests that Fractions A and I contain most of the glycopeptides (combined yield, 80% on the basis of amino acid content). Fraction II is probably a mixture of glycopeptides and peptides*, whereas Fractions III–V contain peptides and corresponding derivatives of amino acids. Fraction VI consisted mainly of the dihydroxy acid amides. This was confirmed by identification, after acid hydrolysis, of glyceric and 2,3-dihydroxybutanoic acid by paper electrophoresis and paper chromatography.

The various serine and threonine contents of the fractions (Fig. 3) led to the conclusion that essentially all these amino acids are glycosylated in native blood-group H glycoprotein. Thus, the content of serine and threonine in Fractions III–VI, which do not contain glycopeptides, is very low, and in Fractions A, I, and II,

*Fraction II consisted of glycopeptides (~90%) which were retained upon ultrafiltration on PSAC Millipore membrane and of peptides (~10%) which passed through it.

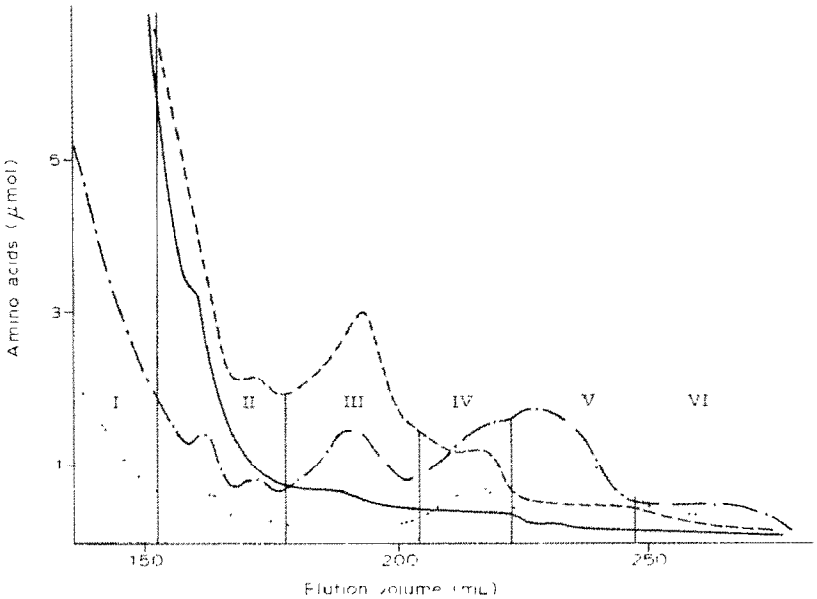


Fig. 3 Gel-filtration pattern of Fractions B plus C (from Fig. 1) on a column (1.4 × 230 cm) of Sephadex G-15 eluted with 0.1M acetic acid and monitored by amino acid content (μmol) after hydrolysis with 6M hydrochloric acid for 24 h at 100° (——) threonine + serine, (------) proline, (– – – – –) alanine, and (··· ··) glycine.

the content of these amino acids is essentially equal to the content of 2-amino-2-deoxy-D-galactose (*cf.* Tables I and II). This sugar may account for a number of carbohydrate chains linked to the peptide backbone, as previous studies have shown that 2-amino-2-deoxy-D-galactose residues are present only in the carbohydrate–peptide linkage of blood-group H glycoprotein[4]

The content of *N*-terminal glyceric acid and 2,3-dihydroxybutanoic acid residues (Table II) is an important characteristic of the glycopeptide and peptide fractions. The ratio of total amino acids to total dihydroxy acids is indicative of the average length of the peptide chain of the products obtained. This ratio was found equal to 11.4, 5.3, and 4.0 for glycopeptide Fractions A, I and II, respectively (Table II).

The ratio of 2-amino-2-deoxy-D-galactose residues (equal to the number of carbohydrate–peptide linkages) to glyceric acid plus 2,3-dihydroxybutanoic acid residues (which is equal to the number of glycopeptide molecules) is a measure of the number of carbohydrate chains in glycopeptides. This ratio was 3.5 for Fraction A, and 1.5 for Fraction I, indicating the presence of several carbohydrate chains. Consequently, the highest-molecular-weight glycopeptides present in Fraction A contain 3–4 carbohydrate chains per peptide chain and those present in Fraction I contain 1–2 carbohydrate chains. Hence, the hydroxyamino acids glycosidically linked to carbohydrate chains in glycopeptides are separated by bridges of 1–3 amino acid residues on the average.

In Fractions III–VI, the ratio of total amino acids to total dihydroxy acids was found to be 2.3, 1.0, 0.5, and 0.25, respectively. Thus, Fraction III consists mainly of *N*-dihydroxyacyldi- and -tri-peptide amides, Fraction IV contains the same derivatives of amino acids, Fraction V is a mixture of *N*-dihydroxyacylamino acids and dihydroxy acid amides, and finally Fraction VI contains dihydroxy acid amides essentially. These data indicate an average size for the peptides in various fractions in agreement with the apparent molecular masses estimated from the elution volumes from Sephadex G-15 (Fig. 2).

In order to determine the distribution of carbohydrate chains on the peptide core of blood-group H glycoprotein, the number of moles of dihydroxy acid amides was compared with the number of moles of dihydroxy acid bound to amino acids and peptides, as estimated from terminal dihydroxy acid content. Determination of the dihydroxy acids content of the low-molecular-weight Fractions III–VI revealed that these acids were present mainly as amides or bound to a single amino acid (Fractions IV–VI). Thus, the carbohydrate chains in blood-group H glycoprotein are rather frequently present in the peptide chain, in blocks, or separated by a single amino acid. Several carbohydrate chains are apparently separated by 2 or 3 amino acids on the average. The content of longer peptides, which may be expected to be present in Fraction II, seems to be insignificant, since ultrafiltration data (filter Millipore PSAC NMWL 10^3) indicated that the molar content of peptides of Fraction II was <2% of those in Fractions III–VI. The average length of the peptide bridge in the glycopeptide fractions that are the principal products of degradation was 1–3 amino acids, whereas, in the peptide fractions, it was 0 or 1 amino acids. Hence, it may be concluded that the rate of β-elimination is unequal for carbohydrate chains located at various sites of the glycoprotein molecule. Carbohydrate chains attached to neighboring hydroxyamino acids or separated by a single amino acid seem to have been eliminated first under the conditions used.

In all fractions (except Fraction VI), the content of dihydroxybutanoic acid was approximately twice that of glyceric acid (Table II), which agrees with the higher content of threonine as compared to serine in blood-group H glycoprotein. The threonine-to-serine ratio in the starting material (2.3:1) and in the high-molecular-weight fraction (2.6:1) is almost the same. Consequently, the elimination of the carbohydrate chains from serine and threonine residues took place at approximately the same rate.

The differences between the amino acid compositions of the fractions were characteristic. The composition of the glycopeptides of Fractions A and I is very close to that of the high-molecular-weight fraction, and the preponderant amino acids are proline, threonine, and serine (Table I). In Fraction III, the proportion of proline was the highest, corresponding to di- and tri-peptide linking-sections, whereas the relative proline content was lower than that of alanine and comparable to that of glycine for linking sections made of a single amino acid (Fig. 3, Fractions IV, V). In the high-molecular-weight fraction, the content of alanine and glycine was 2/5 and 1/4, respectively, that of proline.

The distribution of amino acids in the blood-group H peptide chain presented herein is in good agreement with the structures of glycopeptides reported previously for the blood-group A glycoprotein from human, ovarian-cyst fluid[3] and for the blood-group A and H glycoprotein from pig stomach linings[9]. It is probable that the conclusions drawn from the present work refer only to the hydrophilic part of the peptide chain that is resistant to proteolytic enzymes, because the isolation of the glycoprotein involved autolysis of pig-stomach linings, which may have split off the hydrophobic portion of the molecule.

EXPERIMENTAL

General methods. — Blood-group substance H was isolated, from pig stomach linings having H specificity, as described by Likhosherstov *et al.*[10]. Amino acids and hexosamines were determined with an amino acid analyzer "Biotronik LC 4010", after hydrolysis with 6M hydrochloric acid for 24 h at 100° (for amino acids), and 4M hydrochloric acid for 17 h at 100° (for hexosamines). Neutral monosaccharides were determined with a carbohydrate analyzer "Biotronik LC 2000", after hydrolysis with 4M trifluoroacetic acid for 17 h at 100°. Glyceric acid was quantitatively determined by periodate oxidation, followed by formaldehyde determination according to Vashovsky and Isay[11] with mannitol as a standard. 2,3-Dihydroxybutanoic acid was quantitatively determined by periodate oxidation, followed by acetaldehyde determination by the method of Bhattacharyya and Aminoff[12], as modified by Vidershayn and Kolibaba[13] with 2,3-butanediol as a standard. Paper electrophoresis was performed in 25mM pyridine–acetic acid buffer, pH 4.5, and paper chromatography in 4:1:5 (v/v) 1-butanol–acetic acid–water and 16:1:3 (v/v) ethanol–7M ammonia–water. The compounds were detected with the potassium periodate–silver nitrate–potassium hydroxide (for glycols), aniline–D-xylose (for acids), and chlorine–potassium iodide with starch (for amides) reagents.

Cleavage of N-*acetyldehydroalanine.* — *N*-Acetyldehydroalanine (100 mg) (Reanal, Hungary) was dissolved in water (10 mL), aqueous bromine was added in portions until the solution became yellow, and the pH was adjusted to ~3–4 with 0.2M sodium hydroxide. Additional aqueous bromine (one-half of the original volume) was added and the mixture was kept for 15 min in darkness. The excess of bromine was removed with 20mM sodium thiosulfate, the pH was adjusted to ~9 with M sodium hydroxide, and sodium borohydride was added to a 0.4M concentration, the pH being maintained at 10.8–11.0 for 4 h by addition of 0.1M sodium hydroxide. The solution was made neutral with 50% acetic acid, and passed through a column of AG-50 X2 (H^+, 12 mL) resin. The eluate was concentrated and boric acid was removed from the residue by repeated addition and evaporation of methanol. The bromide ions were removed with silver carbonate, and AG-50 X2 (H^+) resin was added to pH ~4. Glyceric acid and acetamide were identified by comparison with authentic samples by paper electrophoresis and paper chromatography. The yield of glyceric acid was ~90%.

Preparation of the high-molecular-weight fraction from blood-group H glycoprotein. — Blood-group H glycoprotein from pig-stomach linings (2.3 g) was treated[15] with 50mM sodium carbonate for 4 h at 70°. The high-molecular-weight fraction excluded from the Bio-Gel P-100 column (5.0 × 75 cm) was subjected to the same treatment. After repeated fractionation on Bio-Gel P-100, 1.1 g of product was obtained.

Cleavage of the high-molecular-weight fraction at enamine group. — The product just described (1 g) was dissolved in 0.1M acetic acid (20 mL), and aqueous bromine was added portionwise until coloration of an iodide–starch strip. The pH was adjusted to 3–4 with M sodium hydroxide, the solution was concentrated to ~20 mL, and then additional aqueous bromine added (one-half of the original volume) and the mixture kept for 15 min in darkness. The excess of bromine was removed with 20mM sodium thiosulfate, the pH adjusted to ~9 with M sodium hydroxide, sodium borohydride added to a concentration of 0.4M, the pH being kept at 10.8–11.0 for 4 h by addition of M sodium hydroxide, and then the solution was made neutral with 50% acetic acid.

Fractionation of glycopeptides and peptides. — The aforementioned solution was concentrated *in vacuo* and the high-molecular-weight fraction, excluded from a Sephadex G-25 column (3.6 × 120 cm), was collected and lyophilized. The other fractions were concentrated and repeatedly separated on a Sephadex G-25 column (1.8 × 120 cm), into three fractions (A, B, and C; Fig. 1). The high-molecular-weight fraction that was excluded (Fraction A) was lyophilized (yield 0.6 g). Low-molecular-weight Fraction C, which contained much salt, was passed successively through columns of AG-1 X8 (AcO^-; 25 mL) and AG-50 X2 (H^+, 40 mL) resin, and then pooled with Fraction B. These two combined fractions were concentrated and subjected to gel-chromatography on a column (1.4 × 230 cm) of Sephadex G-15 (Fig. 2) in 0.1 M acetic acid. Fractions (3.5 mL) were collected and analyzed for amino acid content (Fig. 3). The eluate was separated into six fractions (I–VI), which were analyzed for amino acids, hexosamines, neutral monosaccharides, and dihydroxy acid residues.

To identify the dihydroxy acids, Fraction VI was hydrolyzed with 2M hydrochloric acid for 3 h at 100°, the solution was evaporated *in vacuo*, water was added to the residue twice and evaporated, traces of hydrochloric acid were removed with silver carbonate, AG-50 X2 (H^+) resin was added to pH ~4, and the solution was concentrated, and analyzed by paper electrophoresis and paper chromatography in the presence of authentic samples.

REFERENCES

1 F. MAISONROUGE-MCAULIFFE AND E. A. KABAT, *Arch. Biochem. Biophys.*, 175 (1976) 90–113.
2 V. A. DEREVITSKAYA, *Pure Appl. Chem.*, 53 (1981) 89–106.
3 S. D. GOODWIN AND W. M. WATKINS, *Eur J. Biochem.*, 47 (1974) 371–382.
4 A. S. R. DONALD, *Biochim. Biophys Acta*, 317 (1973) 420–436.
5 A. PATCHORNIK AND M SOKOLOVSKY, *J. Am. Chem. Soc.*, 86 (1964) 1206–1212.
6 F. DOWNS AND W. PIGMAN, *Biochemistry*, 8 (1969) 1760–1766.

7 V. A. DEREVITSKAYA, S. G KARA-MURZA, AND N. K. KOCHETKOV, *Khim Prir Soedin.*, 4 (1966) 271–277.
8 L. M. LIKHOSHERSTOV, M. D. MARTYNOVA, E. A. KOSHELEVA, V. A. DEREVITSKAYA AND N. K KOCHETKOV, *Dokl. Akad. Nauk SSSR*, 226 (1976) 709–710
9 N. K. KOCHETKOV, V A. DEREVITSKAYA, L M. LIKHOSHERSTOV, AND S A MEDVEDEV, *Dokl. Akad. Nauk SSSR*, 219 (1974) 1498–1501
10 L. M. LIKHOSHERSTOV, V. A DEREVITSKAYA AND V. I. FEDOROVA, *Biokhimiya*, 34 (1969) 45–50
11 V. E. VASKOVSKY AND S. V ISAY, *Anal Biochem.*, 30 (1969) 25–31
12 A K. BHATTACHARYYA AND D. A AMINOFF, *Anal. Biochem* , 14 (1966) 278–283
13 G YA VIDERSHAYN AND L. G. KOLIBABA, *Vopr. Med. Khim.*, 15 (1969) 637–641
14 A. I. USOV AND M. A REKHTER, *Zh. Obshch Khim* , 39 (1969) 912–913.
15 N K. KOCHETKOV, V. A DEREVITSKAYA, L M. LIKHOSHERSTOV, M. D. MARTYNOVA, AND S. N. SENCHENKOVA, *Carbohydr Res.*, 12 (1970) 437–447