Biochem. J. (1981) **193**, 447–457
Printed in Great Britain

Characterization of collagenous and non-collagenous peptides of a glycoprotein isolated from alveoli of patients with alveolar proteinosis

Sambhu N. BHATTACHARYYA
Department of Biochemistry and Department of Medicine, Duke University Medical Center, Durham, NC 27710, U.S.A.

(*Received 17 June 1980/Accepted 17 September 1980*)

A glycoprotein of M_r 62000 was isolated from lung lavage material of patients with alveolar proteinosis. The glycoprotein was found to contain (per molecule) 72 residues of glycine, 5 residues of hydroxyproline, 3 molecules of sialic acid, 4.9 molecules of mannose, 4.0 molecules of galactose, 0.9 molecule of fucose and 7.0 molecules of *N*-acetylglucosamine. Limited pepsin digestion of the glycoprotein resulted in six peptides, three of which contained hydroxyproline and nearly 30% glycine, and two of which contained all the carbohydrate present in the glycoprotein of M_r 62000. The three peptides containing hyroxyproline and with high content of glycine contained a repeating -Gly-X-Y- sequence in the peptide chain. Partial amino acid-sequence anlayses on the peptides derived from the digestion of the alveolar glycoprotein with various proteolytic enzymes indicate that this glycoprotein is characterized by the presence of alternating collagenous and non-collagenous regions in the same polypeptide chain.

Alveolar proteinosis (Rosen *et al.*, 1958) is a chronic pulmonary disease of unknown pathogenesis in which the alveoli and terminal bronchioles of the lung are filled with excessive amounts of periodate–Schiff-positive amorphous material. This material can be removed by pulmonary lavage (Ramirez-R, 1967; Kylstra *et al.*, 1971). After delipidation, the lavage material was analysed by sodium dodecyl sulphate/polyacrylamide-gel electrophoresis (Bhattacharyya *et al.*, 1976*a*; Bhattacharyya & Lynn, 1977*a*) and found to contain three major proteins with M_r 69000, 62000 and 36000. The last two proteins were glycoproteins, and contained hydroxyproline, sialic acid, mannose, galactose, fucose, *N*-acetylglucosamine and a high content of glycine. Two similar glycoproteins have been found in lavage material and lamellar bodies isolated from lungs of different species of normal animals (Bhattacharyya *et al.*, 1975, 1976*b*) and in human amniotic fluid (Bhattacharyya & Lynn, 1978). The structures of the oligosaccharide units of these glycoproteins have been established (Bhattacharyya & Lynn, 1977*a*,*b*).

A soluble glycoprotein of M_r 80000 containing hydroxyproline and with high content of glycine has been isolated from the lung lavage material of patients with alveolar proteinosis and characterized (Bhattacharyya & Lynn, 1979). It has been reported (Bhattacharyya & Lynn, 1977*a*, 1979) that the glycoproteins of M_r 62000 and 36000 were proteolytic products of the glycoprotein of M_r 80000. It has also been reported that the three alveolar glycoproteins contained short collagen-like sequences in the peptide chain, but the precise location of these collagen-like regions in the peptide chain was not known. The present investigation describes the isolation and characterization of collagenous and non-collagenous peptides present in the glycoprotein of M_r 62000.

Experimental

Isolation of glycoprotein from lavage material

Lavage material was obtained from lungs of patients with alveolar proteinosis as described previously (Kylstra *et al.*, 1971; Bhattacharyya *et al.*, 1976*a*). The delipidated lavage material was dissolved in 20mM-Tris/HCl buffer, pH7.5, containing 0.1% sodium dodecyl sulphate and 0.05% 2-mercaptoethanol. The glycoprotein was purified by Bio-Gel P-200 column chromatography as described previously (Bhattacharyya & Lynn, 1977*a*). The purified glycoprotein was dialysed against 4 litres of 100mM-sodium acetate buffer, pH4.5, containing 1M-urea for 48h with two changes. The material was then dialysed in the cold against 4 litres of 100mM-sodium acetate buffer, pH4.5, for 48h with two changes.

Digestion of the glycoprotein of M_r 62000 with pepsin

The purified glycoprotein in 100mM-sodium acetate buffer, pH4.5, was treated with pepsin

(glycoprotein/pepsin ratio 50:1, w/w) at 37°C as described by Reid (1976), except that the period of digestion was 6–8h. A voluminous precipitate was formed during incubation. Under this condition, nearly 85–90% of the glycoprotein was cleaved by pepsin, as judged by the peptide profile seen on a polyacrylamide-gel (10% acrylamide) electrophoretogram. After the incubation, the precipitate was collected by centrifugation and washed three or four times with distilled water. The precipitate was then freeze-dried. The washings and the supernatant were combined, adjusted to pH 8.0 with NaOH and freeze-dried. The freeze-dried material was dissolved in 50mM-pyridine/acetate buffer, pH 4.5, and desalted by chromatography, in the same buffer, on a Bio-Gel P-4 column (1.5cm × 30cm). The fractions containing protein were freeze-dried.

Isolation of peptides from the pepsin-digested material

The freeze-dried pepsin-soluble fraction (supernatant) was dissolved in 10mM-sodium acetate buffer, pH 4.5, containing 2M-urea and was chromatographed on a CM-cellulose column (1.5cm × 25cm) equilibrated with the same buffer. The column was eluted with the equlibrating buffer. before a linear gradient of 0–1.5M-NaCl (200ml) in equilibrating buffer was applied. The fractions containing protein were combined, desalted on a Bio-Gel P-4 column (1.5cm × 30cm) as described above and freeze-dried. The peptides were further purified by chromatography on Sephadex G-50 and G-25 columns (1.5cm × 60cm and 1.5cm × 60cm respectively) (Reid, 1974). The smallest peptides, where necessary, were further purified by paper chromatography and paper electrophoresis, as described previously (Bhattacharyya *et al.*, 1976*a*; Bhattacharyya & Lynn, 1977*a*, 1979).

The freeze-dried precipitate was dissolved in 10mM-sodium phosphate buffer, pH 6.8, containing 2M-urea and applied to a DEAE-cellulose column (1.5cm × 25cm) equilibrated with the same buffer. The column was eluted with the equilibrating buffer before a linear gradient of 0–1.5M-NaCl (200ml) in equilibrating buffer was applied. The fractions containing protein were desalted as described above and freeze-dried. The peptides were further purified by chromatography on Sephadex G-50 and G-25 columns (1.5cm × 60cm and 1.5cm × 60cm respectively) as described previously (Bhattacharyya *et al.*, 1976*a*; Bhattacharyya & Lynn, 1977*a*, 1979).

Digestion of peptide with collagenase

The peptides were partially digested with chromatographically purified collagenase (peptide/collagenase ratio 50:1, w/w) in 50mM-Tris/HCl buffer, pH 7.4, containing 5mM-$CaCl_2$ and 0.25mM-*N*-ethylmaleimide at 37°C for 3h, as described by Reid (1974). The peptides were purified by CM-cellulose column, Sephadex G-25 column and paper chromatography as described above.

Treatment of CNBr-cleavage peptides of the glycoprotein of M_r 62000 with pepsin

Treatment of the glycoprotein of M_r 62000 with CNBr resulted in four peptides of M_r 27000, 18000, 12000 and 5000. Isolation and characterization of these CNBr-cleavage peptides have been reported (Bhattacharyya & Lynn, 1977*a*). Digestion of the CNBr-cleavage peptides with pepsin and purification of pepsin-digest peptides were achieved by CM-cellulose column, DEAE-cellulose column, Sephadex G-25 column and paper chromatography as described above.

Tryptic digestion of the peptides

Various peptides obtained from pepsin-digested, CNBr-cleaved and collagenase-treated materials were treated with trypsin by the procedures described previously (Bhattacharyya & Lynn, 1977*a*, 1979). The peptides were isolated by the same methods as described above.

CNBr treatment of the peptides derived from pepsin digestion and trypsin digestion

The peptides were digested with CNBr by the same procedure as described previously (Bhattacharyya *et al.*, 1976*a*; Bhattacharyya & Lynn, 1977*a*, 1979). In some cases the peptides were reduced and alkylated (Bhattacharyya *et al.*, 1974) before treatment with CNBr. CNBr-cleavage peptides were isolated by Bio-Gel and ion-exchange column chromatography as described previously (Bhattacharyya *et al.*, 1976*a*; Bhattacharyya & Lynn, 1977*a*, 1979). The smallest peptides were further purified by paper chromatography and paper electrophoresis (Bhattacharyya *et al.*, 1976*a*).

Quantitative amino acid analysis, manual dansyl–Edman degradation, partial N-terminal sequence analysis, carbohydrate analysis, electrophoresis and chromatography and digestion with carboxypeptidases A and B

These were performed as described previously (Bhattacharyya & Lynn, 1979).

Sodium dodecyl sulphate / polyacrylamide - gel electrophoresis

This procedure was performed as described previously (Bhattacharyya & Lynn, 1979).

Results

Purification of the peptides obtained after partial digestion of the glycoprotein of M_r 62000 with pepsin

A precipitate was formed when the glycoprotein

of M_r 62000 was digested with pepsin. The precipitate contained carbohydrate, but not hydroxyproline, nor a high content of glycine. The precipitation of these glycopeptides during the incubation may have prevented further degradation of the peptides by pepsin. The acidity of the reaction medium could have contributed to this precipitation, since the precipitate was dissolved when it was suspended in a more basic buffer medium.

The supernatant containing peptides of the alveolar glycoprotein solubilized by digestion with pepsin was fractionated by chromatography on a CM-cellulose column, and four peaks (P1–P3 and P7) were observed (Fig. 1). The peptide eluted at the beginning of the chromatogram was, on the basis of its electrophoretic properties, glycoprotein that had not reacted. The two peaks (P1 and P2) containing two peptides were further purified by Sephadex G-50 column chromatography. Sodium dodecyl sulphate/polyacrylamide-gel electrophoresis indicated a single band from each peak (Figs. 2*c* and 2*d*). The M_r of the peptide in peak P1 was found to be 7000, whereas peak P2 contained a peptide of M_r 5000. When the crude supernatant was electrophoresed in sodium dodecyl sulphate/polyacrylamide gel, two bands of M_r 7000 and 5000, corresponding to peaks P1 and P2, were observed, as shown in Fig. 2(*b*). Fig. 2(*a*) indicated the band of the purified alveolar glycoprotein of M_r 62000. The peptide in peak P3 (Fig. 1) was further purified by Sephadex G-25 column chromatography, and was found to have M_r approx. 2000 by the method described by Bhattacharyya & Lynn (1977*a*, 1979).

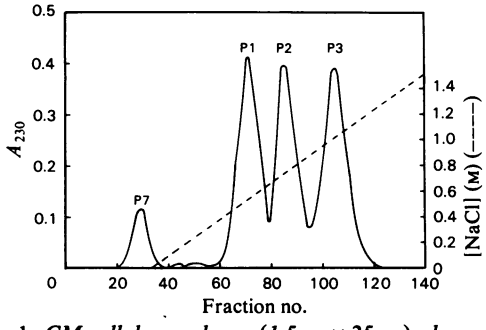

Fig. 1. *CM-cellulose column (1.5 cm × 25 cm) chromatography of the peptides (25 mg) in supernatant obtained after digestion of the glycoprotein of M_r 62000 with pepsin*

The column was equilibrated with 10 mM-sodium acetate buffer, pH 4.5, containing 2 M-urea. After the application of the sample, the column was washed with the equilibrating buffer before elution was performed with a linear gradient of 0–1.5 M-NaCl in 10 mM-sodium acetate buffer, pH 4.5, containing 2 M-urea. The volume per fraction was 1.5 ml.

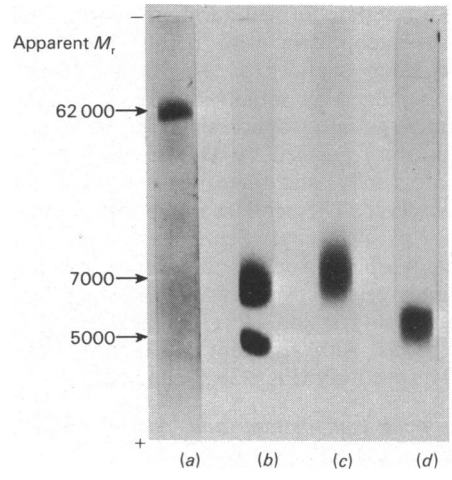

Fig. 2. *Sodium dodecyl sulphate/polyacrylamide-gel electrophoresis of the material in peaks P1 and P2 (Fig. 1)*

Electrophoresis was performed in polyacrylamide gels (10% acrylamide) containing approx. 30 μg of sample in (*a*), 120 μg in (*b*), 75 μg in (*c*) and 60 μg in (*d*). (*a*) Glycoprotein of M_r 62000; (*b*) crude pepsin-solubilized supernatant; (*c*) material from peak P1 (Fig. 1); (*d*) material from peak P2 (Fig. 1). The gels were stained with Coomassie Blue.

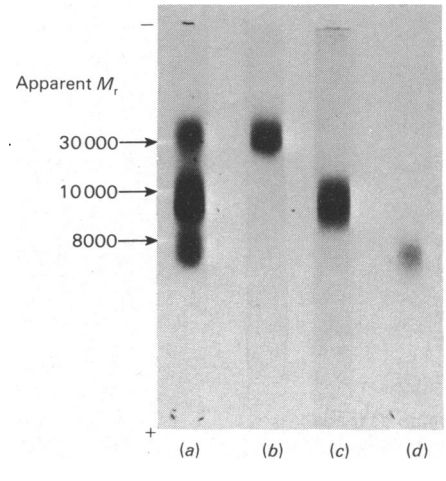

Fig. 3. *Sodium dodecyl sulphate/polyacrylamide-gel electrophoresis of the peptides in the precipitate obtained after digestion of the alveolar glycoprotein with pepsin*

Electrophoresis was performed in polyacrylamide gels (10% acrylamide) containing approx. 150 μg of sample in (*a*), 70 μg in (*b*), 80 μg in (*c*) and 50 μg in (*d*). (*a*) Crude precipitate; (*b*) material from peak P6 (Fig. 5); (*c*) material from peak P4 (Fig. 4); (*d*) material from peak P5 (Fig. 4). The gels were stained with Coomassie Blue.

The precipitate formed during the digestion of the alveolar glycoprotein with pepsin was found to contain three peptides of M_r 30000, 10000 and 8000, as judged by sodium dodecyl sulphate/polyacrylamide-gel electrophoresis (Fig. 3*a*). The three peptides were purified by DEAE-cellulose column chromatography, and three peaks (P4–P6, Fig. 4) were observed. The peptides in the three peaks were further purified by Sephadex G-50 column chromatography. Sodium dodecyl sulphate/polyacrylamide-gel electrophoresis of the purified peptides in each fraction (P4–P6, Fig. 4) resulted in single bands, with apparent M_r 10000, 8000 and 30000 respectively (Figs. 3*c*, 3*d* and 3*b*).

Amino acid composition and partial amino acid sequence of the collagenous peptides

The three peptides (P1–P3, Fig. 1) contained hydroxyproline and 30% glycine, but neither cysteine nor carbohydrate (Table 1). Threonine was absent from peptide P2, whereas histidine was absent from peptides P2 and P3. Methionine was present only in peptide P2. *N*-Terminal amino acid residues for these peptides were threonine, serine and serine respectively, whereas phenylalamine was the *C*-terminal residue in each case.

Partial *N*-terminal amino acid sequence analyses of peptide P1 indicated the sequence Thr-Asx-Gly-Tyr-Ala-Phe-Thr-Gly-Pro-Hyp-Gly-Ser-. Since the same amino acid sequence was found in the *N*-terminal segment of the glycoprotein of M_r 62000 and one of its CNBr-cleavage peptides (Bhattacharyya & Lynn, 1977*a*, 1979), it was concluded that the peptide P1 was the *N*-terminal peptide.

Peptide P3 was partially digested with collagenase, and two peptides (P-Col1 and P-Col2, Fig. 5) were obtained by CM-cellulose column chromatography. The two peptides were further purified by Sephadex G-25 column chromatography. Hydroxyproline was present in peptide

Table 1. *Amino acid and carbohydrate compositions of the purified glycoprotein and its peptides derived from limited digestion with pepsin*

The amino acid composition is expressed as residues per molecule of peptide to the nearest whole number.

Amino acid Peptide ...	P1	P2	P3	P4	P5	P6	Total	Glycoprotein of M_r 62000
4-Hydroxyproline	2	2	1	0	0	0	5	5
Aspartic acid	8	2	1	7	8	27	53	53
Glutamic acid	7	5	1	8	7	31	59	59
Threonine	2	0	1	3	5	11	22	22
Serine	1	2	1	4	3	11	22	22
Proline	4	3	2	6	5	19	39	39
Glycine	19	15	7	4	3	24	72	72
Alanine	3	2	1	8	7	26	47	47
Cysteine*	0	0	0	2	1	5	8	8
Valine	3	3	1	3	2	17	29	29
Methionine	0	1	0	0	1	1	3	3
Isoleucine	2	1	1	6	5	11	26	26
Leucine	4	2	1	9	8	28	52	52
Tyrosine	1	1	1	4	3	5	15	15
Phenylalanine	2	2	1	4	3	16	28	28
Lysine	2	2	1	5	3	12	25	25
Histidine	1	0	0	3	3	7	14	14
Arginine	2	2	1	7	4	14	30	30
Total residues	63	45	22	83	71	265	549	549
M_r by amino acid analysis	6930	4950	2420	9130	7810	29150	60390	
M_r by gel electrophoresis and Bio-Gel chromatography	7000	5000	2000	10000	8000	30000	62000	
N-Terminus	Thr	Ser	Ser	Ser	Ala	Ser		Thr
Yield (%)	80	78	65	70	65	75		
Carbohydrate content (mol/mol)								
Sialic acid	0	0	0	1.5	0.8	0		3.0
Mannose	0	0	0	2.4	2.3	0		4.9
Galactose	0	0	0	2.0	2.1	0		4.0
Fucose	0	0	0	0.8	0	0		0.9
Glucosamine	0	0	0	3.5	3.5	0		7.0

* Measured as cysteic acid.

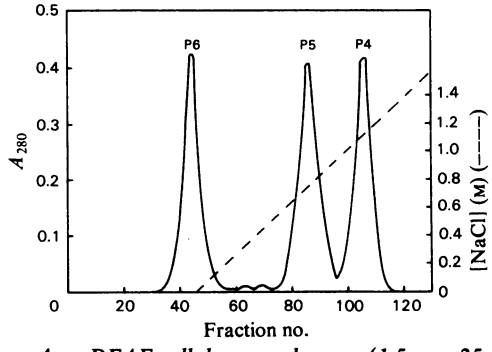

Fig. 4. *DEAE-cellulose column (1.5 cm × 25 cm) chromatography of the peptides (25 mg) in the precipitate obtained after digestion of the alveolar glycoprotein with pepsin (see details in the text)*

The column was equilibrated with 10mM-sodium phosphate buffer, pH 7.2, containing 2M-urea. After the application of the sample, the column was washed with the equilibrating buffer before a linear gradient of 0–1.5M-NaCl in 10mM-sodium phosphate buffer, pH 7.2, containing 2M-urea was applied. The volume per fraction was 1.4ml.

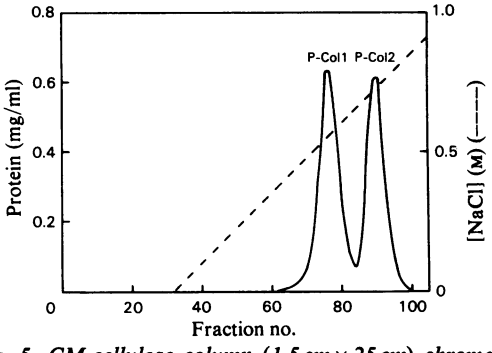

Fig. 5. *CM-cellulose column (1.5 cm × 25 cm) chromatography of the peptides (5 mg) derived from the digestion of peptide P3 (Fig. 1) with collagenase*

The column was equilibrated with 10mM-sodium acetate buffer, pH 4.5, containing 2M-urea. After the application of the sample, the column was washed with the equilibrating buffer before elution was performed with a linear gradient of 0–1.0M-NaCl in 10mM-sodium acetate buffer, pH 4.5, containing 2M-urea. The volume per fraction was approx. 1.0ml.

P-Col2, whereas peptide P-Col1 did not contain this amino acid (Table 2). Both peptides contained 30% glycine. Glycine and serine were the *N*-terminal residues of the peptides P-Col1 and P-Col2 respectively, whereas the *C*-terminal residues were phenylalanine and hydroxyproline.

Table 2. *Amino acid composition of the peptides obtained after treatment of purified pepsin-digest peptide with collagenase*

The amino acid composition is expressed as residues per molecule of peptide.

	Peptide P3	
Amino acid Peptide ...	P-Col1	P-Col2
4-Hydroxyproline	0	1
Aspartic acid	1	0
Glutamic acid	0	1
Threonine	1	0
Serine	0	1
Proline	1	1
Glycine	4	3
Alanine	0	1
Cysteine*	0	0
Valine	1	0
Methionine	0	0
Isoleucine	1	0
Leucine	0	1
Tyrosine	1	0
Phenylalanine	1	0
Lysine	1	0
Histidine	0	0
Arginine	0	1
Total residues	12	10
N-Terminus	Gly	Ser

* Measured as cysteic acid.

Amino acid and carbohydrate composition of the non-collagenous peptides

Digestion of the alveolar glycoprotein with pepsin resulted in a precipitate containing three non-collagenous peptides (P4, P5 and P6, Fig. 4). The three peptides contained cysteine, a low amount of glycine and no hydroxyproline (Table 1). Peptide P4 did not contain any methionine. Peptides P4 and P5 accounted for all the carbohydrate present in the native glycoprotein. No carbohydrate was present in peptide P6. Serine, alanine and serine were the *N*-terminal residues of peptides P4, P5 and P6 respectively, whereas serine and phenylalanine were the *C*-terminal residues of peptides P4 and P5 respectively.

Amino acid and carbohydrate composition of the peptides obtained after digestion of CNBr-cleavage peptides with pepsin

Isolation and characterization of the CNBr-cleavage peptides of M_r 62000 have been reported previously (Bhattacharyya & Lynn, 1977*a*).

Treatment of the CNBr-cleavage peptide of M_r 18000 with pepsin and fractionation of the digested material by CM-cellulose column, DEAE-cellulose

column, Sephadex G-25 column and paper chromatography resulted in three peptides, CP1, CP2 and CP3, of M_r 7000, 10000 and 1000 respectively. Peptide CP1 contained hydroxyproline, 30% glycine and no carbohydrate (Table 3). Cysteine and methionine were absent from this peptide. The amino acid composition and *N*-terminal and *C*-terminal residues of peptide CP1 were identical with those of peptide P1 (Table 1). Partial *N*-terminal amino acid sequence analyses of peptide CP1 resulted in Thr-Asx-Gly-Ala-Phe-Thr-Gly-Pro-Hyp-Gly-Ser-, indicating that peptide CP1, like peptide P1, was the *N*-terminal peptide of the native glycoprotein molecule. Peptide CP2 did not contain any hydroxyproline, nor a high content of glycine, but contained carbohydrate (Table 3). The overall composition and *N*-terminal and *C*-terminal residues of peptide CP2 were identical with those of peptide P4 (Table 1). Hydroxyproline and carbohydrate were not detected in peptide CP3 (Table 3). Amino acid-sequence analyses on this peptide, purified by paper chromatography (R_F 0.70), resulted in Ala-Pro-Gly-Ser-Val-Lys-Thr-Glx-Met.

Treatment of the CNBr-cleavage peptide of M_r 12000 with pepsin and fractionation of the peptides by DEAE-cellulose column and Sephadex G-25 column chromatography resulted in two peptides, CP4 and CP5, of M_r approx. 7000 and 5000 respectively. Peptide CP4 did not contain any hydroxyproline, nor a high content of glycine (Table 3). Methionine was absent from this peptide. The carbohydrate composition of this peptide was similar to that of peptide P5 (Table 1). The overall amino acid composition of peptide CP4 indicated that, except for nine amino acid residues (Table 3), this peptide was identical with peptide P5 (Table 1). Peptide CP5 contained hydroxyproline, approx. 30% glycine and no carbohydrate (Table 3).

Table 3. *Amino acid and carbohydrate compositions of the purified peptides and glycopeptides obtained after treatment of CNBr-cleavage peptides with pepsin*

The amino acid composition is expressed as residues per molecule of peptide.

M_r of CNBr-cleavage peptide ...		18000			12000		5000		
Amino acid	Peptide ...	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8
4-Hydroxyproline		2	0	0	0	2	1	0	0
Aspartic acid		8	7	0	8	2	1	3	0
Glutamic acid		7	8	1	6	4	1	3	1
Threonine		2	3	1	4	0	1	2	0
Serine		1	4	1	2	2	1	1	0
Proline		4	6	1	4	3	2	0	0
Glycine		19	4	1	2	14	7	0	1
Alanine		3	8	1	6	2	1	3	0
Cysteine*		0	2	0	1	0	0	1	0
Valine		3	3	1	1	3	1	2	0
Methionine†		0	0	1	0	1	0	1	0
Isoleucine		2	6	0	5	1	1	1	0
Leucine		4	9	0	8	2	1	3	0
Tyrosine		1	4	0	3	1	1	0	0
Phenylalanine		2	4	0	3	1	1	1	1
Lysine		2	5	1	2	2	1	1	0
Histidine		1	3	0	3	0	0	1	0
Arginine		2	7	0	4	2	1	1	0
Total residues		63	83	9	62	42	22	24	3
N-Terminus		Thr	Ser	Ala	Phe	Ser	Ser	Ser	Glx
Carbohydrate content (mol/mol)									
Sialic acid		0	1.6	0	0.8	0	0	0	0
Mannose		0	2.3	0	2.4	0	0	0	0
Galactose		0	2.0	0	2.0	0	0	0	0
Fucose		0	0.9	0	0	0	0	0	0
Glucosamine		0	3.5	0	3.5	0	0	0	0

* Determined as cysteic acid.
† Determined as homoserine.

Threonine, cysteine and histidine were absent from peptide CP5. Except for three amino acid residues, the amino acid composition of this peptide was identical with that of peptide P2 (Table 1).

Digestion of the CNBr-cleavage peptide of M_r 5000 with pepsin and separation of the peptides by CM-cellulose column, Sephadex G-25 column and paper chromatography resulted in three peptides, CP6, CP7 and CP8, which did not contain any carbohydrate (Table 3). Peptide CP6 contained hydroxyproline and 30% glycine. Cysteine, methionine and histidine were absent from this peptide. The amino acid composition and *N*-terminal and *C*-terminal residues of peptide CP6 (Table 3) were identical with those of peptide P3 (Table 1). Collagenase treatment of the peptide CP6 resulted in two peptides, P-Col1 and P-Col2 (results not shown), the overall compositions of which were identical with those of the same peptides (Table 2) isolated from peptide P3 (Table 1). Peptide CP7 contained no hydroxyproline, proline, glycine or tyrosine (Table 3). Carbohydrate was not detected in this peptide. Serine was the *N*-terminal residue of peptide CP7 (Table 3). Peptide CP8 was isolated by paper chromatography (R_F 0.35) and contained three amino acids (Table 3). The amino acid sequence of this peptide was Glx-Gly-Phe.

Amino acid composition of the peptides treated with trypsin

Peptide P6 (Fig. 1 and Table 1) was digested with trypsin, and two peptides of approx. M_r 26000 and 4000, corresponding to peptides TP1 and TP2, were isolated by Sephadex G-75 column chromatography. Peptide TP1 did not contain hydroxyproline, nor any methionine (Table 4). Carbohydrate was absent from this peptide, and alanine was the *N*-terminal residue. The overall composition of peptide TP1 (Table 4) indicated that this peptide was identical with the peptide of M_r 26000 isolated from the glycoprotein of M_r 62000 with trypsin (Bhattacharyya & Lynn, 1977*a*). Limited *N*-terminal amino acid-sequence analyses of peptide TP1 indicated the sequence Ala-Ser-Pro-Tyr-Gly-, which was identical with the same sequence in the peptide of M_r 26000 (Bhattacharyya & Lynn, 1977*a*). Thus, like the peptide of M_r 26000 (Bhattacharyya & Lynn, 1977*a*), peptide TP1 was the *C*-terminal peptide of the glycoprotein of M_r 62000. Peptide TP2 contained neither hydroxyproline nor tyrosine (Table 4). The *N*-terminal amino acid residue of this peptide was serine, whereas arginine was the *C*-terminal residue. Thus trypsin cleaved the -Arg-Ala- bond of the peptide P6 to yield peptides TP1 and TP2.

Table 4. *Amino acid composition of the peptides obtained after treatment of purified pepsin-digested CNBr-cleavage and collagenase-digest peptides with trypsin*

The amino acid composition is expressed as residues per molecule of peptide.

	Peptide ...	P6		CNBr 27000†		P-Col1		P-Col2	
Amino acid	Peptide ...	TP1	TP2	TC1	TC2	T1-PCol1	T2-PCol1	T1-PCol1	T2-PCol2
4-Hydroxyproline		0	0	0	0	0	0	0	1
Aspartic acid		23	4	23	1	1	0	0	0
Glutamic acid		27	4	27	1	0	0	1	0
Threonine		9	2	9	0	0	1	0	0
Serine		9	2	9	1	0	0	1	0
Proline		18	1	18	1	1	0	0	1
Glycine		22	2	22	2	3	1	2	1
Alanine		22	4	22	1	0	0	1	0
Cysteine*		4	1	4	0	0	0	0	0
Valine		15	2	15	0	0	1	0	0
Methionine		0	1	0	0	0	0	0	0
Isoleucine		10	1	10	0	1	0	0	0
Leucine		24	4	24	1	0	0	1	0
Tyrosine		5	0	5	0	1	0	0	0
Phenylalanine		15	1	15	0	0	1	0	0
Lysine		11	1	11	0	1	0	0	0
Histidine		6	1	6	0	0	0	0	0
Arginine		12	2	12	1	0	0	1	0
Total residues		232	33	232	9	8	4	7	3
N-Terminus		Ala	Ser	Ala	Ser	Gly	Val	Ser	Gly

* Determined as cysteic acid.

† This CNBr-cleavage peptide was obtained as described previously (Bhattacharyya & Lynn, 1977*a*).

When the CNBr-cleavage peptide of M_r 27000 isolated from the glycoprotein of M_r 62000 (Bhattacharyya & Lynn, 1977*a*) was treated with trypsin, two peptides, TC1 and TC2, were obtained by Sephadex G-75 and G-25 column and paper chromatography. Hydroxyproline and methionine were absent from peptide TC1 (Table 4). The overall composition and *N*-terminal residue of peptide TC1 were identical with those of peptide TP1 (Table 4) as well as of the peptide of M_r 26000 (Bhattacharyya & Lynn, 1977*a*). Nine amino acid residues were found to be present in peptide TC2 (Table 4) purified by paper chromatography (R_F 0.33). Amino acid-sequence analyses of this peptide resulted in a sequence Ser-Glx-Gly-Pro-Ala-Asx-Leu-Gly-Arg, which was identical with the sequence of the peptide of M_r 1000 isolated previously from the glycoprotein of M_r 62000 (Bhattacharyya & Lynn, 1977*a*). Trypsin cleaved the -Arg-Ala- bond of the CNBr-cleavage peptide of M_r 27000 to yield peptides TC1 and TC2.

Treatment of peptide P-Col1 (Fig. 5 and Table 2) with trypsin resulted in two peptides. T1-PCol1 and T2-PCol1, which were isolated by Sephadex G-10 column and paper chromatography. Peptide T1-PCol1, purified by paper chromatography (R_F 0.44), was found to contain eight amino acid residues (Table 4), and amino acid-sequence analyses of this peptide gave the sequence of Gly-Asx-Ile-Gly-Tyr-Pro-Gly-Lys. The amino acid sequence of peptide T1-PCol1 isolated paper chromatography (R_F 0.20) was found to be Val-Gly-Thr-Phe. Thus the two peptides, T1-PCol1 and T2-PCol1, were formed by the cleavage of the -Lys-Val- bond of peptide P-Col1 by trypsin.

Peptide P-Col2 (Fig. 5 and Table 2) was treated with trypsin, and two peptides, T1-PCol2 and T2-PCol2, were isolated by Sephadex G-10 column and paper chromatography. The amino acid sequence of peptide T1-PCol1, purified by paper chromatography (R_F 0.38), was determined (Table 4) and found to be Ser-Gly-Glx-Leu-Gly-Ala-Arg.

Table 5. *Amino acid and carbohydrate compositions of the purified peptides and glycopeptides obtained after treatment of purified pepsin-digest peptides with CNBr*

The amino acid composition is expressed as residues per molecule of peptide.

	Peptide ...	P2		P5		P6		TP2	
Amino acid	Peptide ...	PC1	PC2	PC3	PC4	PC5	PC6	TPC1	TPC2
4-Hydroxyproline		2	0	0	0	0	0	0	0
Aspartic acid		2	0	8	0	24	3	3	1
Glutamic acid		4	1	6	1	28	3	3	1
Threonine		0	0	4	1	9	2	2	0
Serine		2	0	2	1	10	1	1	1
Proline		3	0	4	1	19	0	0	1
Glycine		14	1	2	1	24	0	0	2
Alanine		2	0	6	1	23	3	3	1
Cysteine*		0	0	1	0	4	1	1	0
Valine		3	0	1	1	15	2	2	0
Methionine†		1	0	0	1	0	1	1	0
Isoleucine		1	0	5	0	10	1	1	0
Leucine		2	0	8	0	25	3	3	1
Tyrosine		1	0	3	0	5	0	0	0
Phenylalanine		1	1	3	0	15	1	1	0
Lysine		2	0	2	1	11	1	1	0
Histidine		0	0	3	0	6	1	1	0
Arginine		2	0	4	0	13	1	1	1
Total residues		42	3	62	9	241	24	24	9
N-Terminus		Ser	Glx	Phe	Ala	Ser	Ser	Ser	Ser
Carbohydrate content (mol/mol)									
Sialic acid		0	0	0.7	0	0	0	0	0
Mannose		0	0	2.3	0	0	0	0	0
Galactose		0	0	2.0	0	0	0	0	0
Fucose		0	0	0	0	0	0	0	0
Glucosamine		0	0	3.5	0	0	0	0	0

* Measured as *S*-carboxymethylcysteine.

† Measured as homoserine.

Peptide T2-PCol2 was purified by paper chromatography (R_F 0·30) and found to contain three amino acid residues (Table 4). The amino acid sequence of this peptide was found to be Gly-Pro-Hyp. Trypsin cleaved the -Arg-Gly- bond of peptide P-Col2 to yield peptides T1-PCol2 and T2-PCol2. The complete amino acid sequence of peptide P3 (Table 1) is shown in Fig. 6.

CNBr treatment of the peptides from pepsin-digested glycoprotein of M_r 62000

Studies were performed to learn more about the peptide backbone of the alveolar glycoprotein by subjecting the pepsin-digested peptides (Table 1) containing methionine to CNBr treatment. Peptide P2 (Table 1) was treated with CNBr, and two peptides, PC1 and PC2, were isolated by Sephadex G-25 column and paper chromatography. Peptide PC1 contained two residues of hydroxyproline, 30% glycine and no carbohydrate (Table 5). Threonine, cysteine and histidine were absent from this peptide. The amino acid composition and *N*-terminal and *C*-terminal residues indicated that peptide PC1 was identical with peptide CP5 (Table 3). Peptide PC2 (Table 5) was separated by paper chromatography (R_F 0.35) and contained the same amino acid sequence, Glx-Gly-Phe, as that of peptide CP8 (Table 3).

Peptide P5 (Table 1) was reduced and alkylated before treatment with CNBr. Two peptides, PC3 and PC4, were obtained by DEAE-cellulose column, Sephadex G-25 column and paper chromatography. Peptide PC3 did not contain any hydroxyproline, nor a high content of glycine, but contained

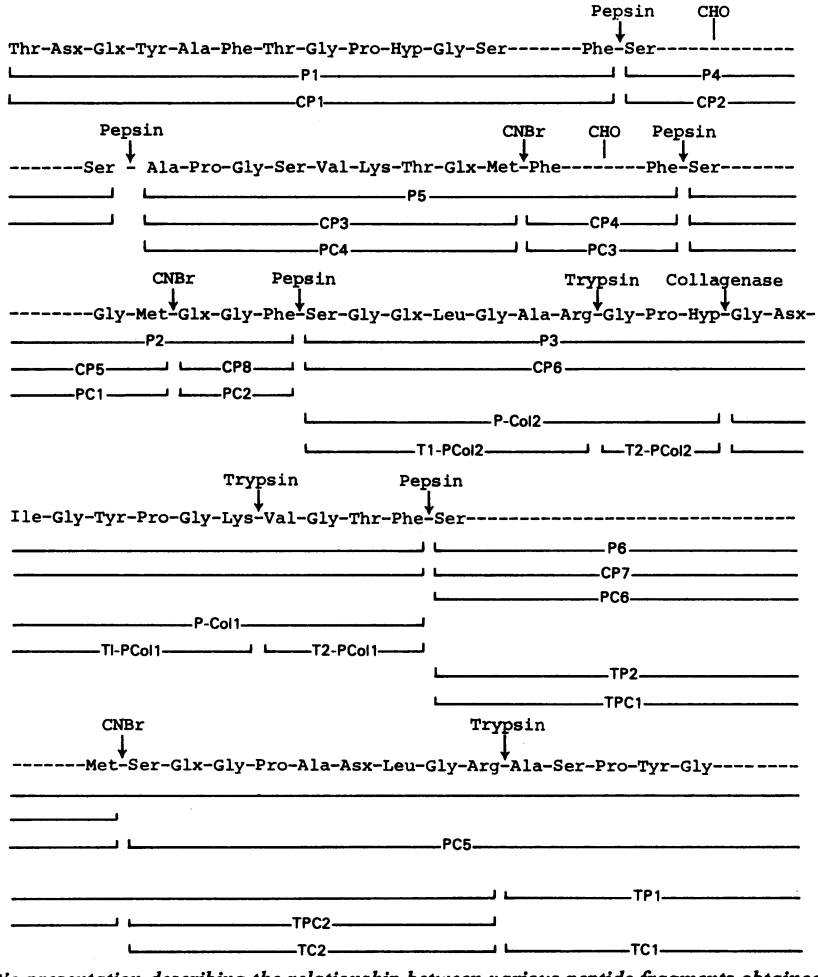

Fig. 6. *Schematic presentation describing the relationship between various peptide fragments obtained by the digestion of the alveolar glycoprotein of M_r 62000 with pepsin, collagenase, trypsin and CNBr*
See the text, Tables and Figures for full details. Abbreviation: CHO, carbohydrate.

carbohydrate (Table 5). This peptide was identical with peptide CP4 (Table 3). Peptide PC4 was purified by paper chromatography (R_F 0.70), and did not contain a high content of glycine or hydroxyproline, nor any carbohydrate (Table 5). The amino acid sequence of this peptide was identical with that of peptide CP3 (Table 3), which was Ala-Pro-Gly-Ser-Val-Lys-Thr-Glx-Met.

CNBr treatment of peptide P6 (Table 1) resulted in two peptides, PC5 and PC6, which were separated by Sephadex G-50 column chromatography. Peptide PC5 did not contain hydroxyproline or methionine (Table 5). Peptide PC5 was identical with the CNBr-cleavage peptide of M_r 27000 (Bhattacharyya & Lynn, 1977*a*), and, except for nine amino acid residues, this peptide (Table 5) was identical with peptides TP1 and TC1 (Table 4). Peptide PC6 did not contain any hydroxyproline, proline, glycine or tyrosine (Table 5), and, except for nine amino acid residues, this peptide was identical with peptide TP2 (Table 4).

Peptide TP2 (Table 4) was treated with CNBr, and two peptides, TPC1 and TPC2, were obtained by Sephadex G-10 column and paper chromatography. The amino acid composition and *N*-terminal and *C*-terminal residues of peptide TPC1 were identical with those of peptide PC6 (Table 5). The amino acid sequence of peptide TPC2 (Table 5), purified by paper chromatography (R_F 0.33), was Ser-Gly-Pro-Ala-Asx-Leu-Gly-Arg. This sequence was identical with that of peptide TC2 (Table 4), as well as with that of the peptide of M_r 1000 isolated previously from the glycoprotein of M_r 62000 (Bhattacharyya & Lynn, 1977*a*).

A schematic presentation of the relationship between different peptides of the alveolar glycoprotein of M_r 62000 obtained from digestion with pepsin, CNBr, collagenase and trypsin is shown in Fig. 6.

Discussion

A glycoprotein of M_r 62000 containing hydroxyproline and a high content of glycine has been isolated from lung lavage of patients with alveolar proteinosis. This glycoprotein with similar composition has been found to be present in the lung lavage and lamellar bodies of normal animals (Bhattacharyya *et al.*, 1975, 1976*b*). It has been reported previously (Bhattacharyya & Lynn, 1977*a,b*) that the glycoprotein contained asparagine-linked carbohydrate moieties as well as a -Gly-Pro-Hyp-Gly- sequence in the same polypeptide chain. Also, the glycoprotein of M_r 62000 was found to be a proteolytic product of the glycoprotein of M_r 80000 (Bhattacharyya & Lynn, 1979). Both the *N*-terminal and *C*-terminal segments of the glycoproteins of M_r 80000 and 62000 contained no hydroxyproline, nor a high amount of glycine (Bhattacharyya & Lynn, 1977*a*, 1979). The present study confirms previous observations (Bhattacharyya & Lynn, 1977*a*, 1979) that the glycoprotein of M_r 36000 is a proteolytic product of the glycoprotein of M_r 62000.

Under controlled condition, pepsin digestion of the alveolar glycoprotein resulted in three collagenous peptides containing a repeating -Gly-X-Y- sequence in the peptide chain. Such triplet sequences have been observed in all types of collagens (Hannig & Nordwig, 1967), including human complement subcomponent Clq (Reid, 1974, 1976, 1979). Basement-membrane or type-IV collagens and acetylcholinesterase from different sources have also been found to contain collagenous and non-collagenous segments in the same polypeptide chain (Spiro, 1967; Kefalides, 1970; Hudson & Spiro, 1972; Sato & Spiro, 1976; Rosenberry & Richardson, 1977; Furuto & Miller, 1980; Crouch *et al.*, 1980; Schwartz *et al.*, 1980). However, the alveolar glycoprotein is very different from collagens, complement subcomponent Clq or basement-membrane glycoproteins in that this glycoprotein lacks hydroxylysine or hydroxylysine-linked carbohydrate moieties in the peptide chain. Instead, the alveolar glycoprotein contains asparagine-linked carbohydrate moieties in the peptide chain (Bhattacharyya *et al.*, 1976*a*; Bhattacharyya & Lynn, 1977*a,b*). Also, the alveolar glycoprotein contains much smaller collagenous sequences than those of the collagens (Hannig & Nordwig, 1967). In this respect, the collagenous sequences in the alveolar glycoproteins resemble those of complement subcomponent Clq (Reid, 1974, 1976, 1979). But the alveolar glycoprotein differs from complement subcomponent Clq in that the former does not contain any hydroxylysine, and the molecular size of the former is much higher than those of the A-chain, B-chain and C-chain subunits of complement subcomponent Clq (Reid, 1974, 1976). Also, the antibody raised against the alveolar glycoprotein (Bhattacharyya & Lynn, 1978) did not react with complement subcomponent Clq (results not shown). It is noteworthy that the presence of asparagine-linked and hydroxylysine-linked oligosaccharides as well as the amino acid sequence -Gly-Pro-Hyp- in a glycoprotein from glomerular basement membrane has been reported by Levine & Spiro (1979). It cannot be completely ruled out, however, that the alveolar glycoproteins of M_r 62000 and 80000 may be proteolytic products of a much larger glycoprotein molecule that, like basement-membrane glycoproteins, contains hydroxylysine and hydroxylysine-linked carbohydrate. Both the *N*-terminal and *C*-terminal propeptides of procollagen have been found to contain mannose and glucosamine (Clark & Kefalides, 1976, 1978), though the sites of attachment have not yet been established.

The unusual feature of this glycoprotein is that the *N*-terminal segment is represented by a short non-collagen-like region followed by a repeating -Gly-X-Y- sequence, which is then followed by oligosaccharide-containing peptide regions (Fig. 6). These regions are then followed by a short collagen-like region. Thus, in contrast with collagens and procollagens, the alveolar glycoprotein is characterized by alternating short collagenous and non-collagenous segments, as opposed to a long collagenous peptide that is either preceded or followed by non-collagenous regions. The presence of collagenous and non-collagenous domains in the polypeptide chains of human complement subcomponent Clq has also been reported by Reid (1974, 1976, 1979). Thus the alveolar glycoprotein, like basement-membrane collagen or complement subcomponent Clq, may represent an entirely different class of collagen-like glycoproteins present in the animal system.

This investigation was supported by National Heart and Lung Institute Grant HL 21480. I thank Dr. William S. Lynn for his advice throughout this work. I acknowledge the excellent technical assistance of Ms. Mary Lynn. I appreciate the assistance of Ms. Trudy Dameron and Ms. Betty Watkins in preparing the manuscript for publication.

References

Bhattacharyya, S. N. & Lynn, W. S. (1977*a*) *Biochim. Biophys. Acta* **494**, 150–161

Bhattacharyya, S. N. & Lynn, W. S. (1977*b*) *J. Biol. Chem.* **252**, 1172–1180

Bhattacharyya, S. N. & Lynn, W. S. (1978) *Biochim. Biophys. Acta* **537**, 329–335

Bhattacharyya, S. N. & Lynn, W. S. (1979) *J. Biol. Chem.* **254**, 5191–5198

Bhattacharyya, S. N., Passero, M. A. & Lynn, W. S. (1974) *Biochim. Biophys. Acta* **342**, 343–350

Bhattacharyya, S. N., Passero, M. A., DiAugustine, R. P. & Lynn, W. S. (1975) *J. Clin. Invest.* **55**, 914–920

Bhattacharyya, S. N., Sahu, S. & Lynn, W. S. (1976*a*) *Biochim. Biophys. Acta* **427**, 91–106

Bhattacharyya, S. N., Rose, M. C., Lynn, M. G., McLeod, C., Alberts, M. & Lynn, W. S. (1976*b*) *Am. Rev. Respir. Dis.* **114**, 843–850

Clark, C. C. & Kefalides, N. A. (1976) *Proc. Natl. Acad. Sci. U.S.A.* **73**, 34–38

Clark, C. C. & Kefalides, N. A. (1978) *J. Biol. Chem.* **253**, 47–51

Crouch, E., Sage, H. & Bornstein, P. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 745–749

Furuto, D. K. & Miller, E. J. (1980) *J. Biol. Chem.* **255**, 290–295

Hannig, K. & Nordwig, A. (1967) in *Treatise on Collagen* (Ramachandran, G. A., ed.), vol. 1, pp. 73–101, Academic Press, New York and London

Hudson, B. G. & Spiro, R. G. (1972) *J. Biol. Chem.* **247**, 4239–4247

Kefalides, N. A. (1970) in *Chemistry and Molecular Biology of the Intracellular Matrix* (Balazs, E. A., ed.), vol. 1, pp. 535–573, Academic Press, New York and London

Kylstra, J. A., Rausch, D. C., Hall, D. K. & Spock, A. (1971) *Am. Rev. Respir. Dis.* **103**, 651–665

Levine, M. J. & Spiro, R. G. (1979) *J. Biol. Chem.* **254**, 8121–8124

Ramirez-R, J. (1967) *Arch. Intern. Med.* **119**, 147–156

Reid, K. B. M. (1974) *Biochem. J.* **141**, 189–203

Reid, K. B. M. (1976) *Biochem. J.* **155**, 5–17

Reid, K. B. M. (1979) *Biochem. J.* **179**, 367–371

Rosen, S. H., Castleman, B. & Liebow, E. A. (1958) *N. Engl. J. Med.* **258**, 1123–1142

Rosenberry, T. L. & Richardson, J. M. (1977) *Biochemistry* **15**, 3550–3558

Sato, T. & Spiro, R. G. (1976) *J. Biol. Chem.* **251**, 4062–4070

Schwartz, D., Chin-Quee, T. & Veis, A. (1980) *Eur. J. Biochem.* **103**, 21–27

Spiro, R. G. (1967) *J. Biol. Chem.* **242**, 1923–1932