Metafora, S., Terada, M., Dow, L. W., Marks, P. A., and Bank, A. (1972), *Proc. Nat. Acad. Sci. U. S. 69*, 1299.
Miller, A. O. A. (1968), *Biochem. Biophys. Res. Commun. 30*, 267.
Morgan, R. S. (1966), *Biochim. Biophys. Acta 123*, 623.
Shapira, G., Rosa, J., Maleknia, N., and Padieu, P. (1968), *Methods Enzymol. 12*, 747.
Wacker, A., Ishimoto, M., and Chandra, P. (1967), *Z. Naturforsch. 22B*, 413.
Walter, G., Zillig, W., Palm, P., and Fuchs, E. (1967), *Eur. J. Biochem. 3*, 194.
Weller, D. L., and Morgan, R. S. (1967), *Biochemistry 6*, 983.
Zoellner, N., and Fellig, J. (1953), *Amer. J. Physiol. 173*, 223.

Structure of α_1-Acid Glycoprotein. The Complete Amino Acid Sequence, Multiple Amino Acid Substitutions, and Homology with the Immunoglobulins†

Karl Schmid,* Hugo Kaufmann, Satoko Isemura, Francine Bauer, Junji Emura, Toshiko Motoyama, Masatsune Ishiguro, and Shigeru Nanno

ABSTRACT: The linear amino acid sequence of the amino-terminal CNBr fragment of α_1-acid glycoprotein derived from pooled human plasma was elucidated and proved to consist of 111 residues. For this investigation the amino acid sequences of the peptides and glycopeptides of a chymotryptic and a tryptic digest and some of the peptides and glycopeptides of a peptic hydrolysate of this protein were elucidated. These data together with the amino acid sequence of the carboxyl-terminal CNBr fragment reported earlier completely established the amino acid sequence of α_1-acid glycoprotein. The five heteropolysaccharide groups of this protein were demonstrated to be linked N-glycosidically to asparaginyl residues. The number of amino acids between two subsequent carbohydrate units differs considerably. This report is thus the first one in which the sequence of a glycoprotein with such a high number of polysaccharide units is described. The following two findings were very unusual. First, 11 amino acid substitutions were detected. The carboxyl-terminal CNBr fragment possesses ten further amino acid replacements as described earlier, so that in 21 of the 181 residues of the protein, or 12%, such substitutions have occurred. These replacements, except for two, can be explained by single point mutations. Secondly, a significant degree of homology was noted between the amino-terminal 43-residue segment of CNBr-I and the amino terminal of the variable region of the κ-type L chain of human IgG. This homology and that of the carboxyl-terminal region of this glycoprotein with the constant region of the H chain of IgG suggest that α_1-acid glycoprotein may represent a protein that is related to the ancestral immunoglobulin.

In a recent publication (Ikenaka *et al.*, 1972) we have described the amino acid sequence of the carboxyl-terminal cyanogen bromide (CNBr)[1] fragment of α_1 acid glycoprotein. In the present paper the amino acid sequence of the chymotryptic, tryptic, and certain peptic peptides and glycopeptides of the remaining part of the polypeptide chain, the amino-terminal CNBr fragment (Ikenaka *et al.*, 1972) of this human plasma protein, is presented.

Materials and Methods

α_1-Acid glycoprotein was isolated from Cohn fraction VI of pooled normal human plasma by a procedure described earlier (Bürgi and Schmid, 1961). This globulin, which possesses a single polypeptide chain (Ikenaka *et al.*, 1966), was demonstrated to be homogeneous as judged by several criteria of purity (Ikenaka *et al.*, 1966; Jeanloz, 1972). Desialyzation was accomplished by mild acid hydrolysis at pH 1.6 and 80° for 1 hr using a 2% protein solution (Schmid *et al.*, 1967). In previous investigations it was shown that this procedure does not lead to the formation of new amino-terminal amino acids (Ikenaka *et al.*, 1966). After dialysis against cold 1% pyridine, the resulting protein solution was lyophilized and used for part of the present study.

The preparation and characterization of the amino-terminal CNBr fragment (CNBr-I) of native α_1-acid glycoprotein, the enzymes, and various other reagents used in the present study were described earlier (Ikenaka *et al.*, 1972). In addition, Nagarse and Pronase (70,000 p.u.k/g) were products from Kaken & Co., Tokyo, while pyrrolidonecarboxylyl-peptidase was prepared essentially according to Doolittle (1972). The specific activity of the enzyme preparation obtained after the Sephadex G-200 step amounted to 3000 units/

† From the Department of Biochemistry, Boston University School of Medicine, Boston University Medical Center, Boston, Massachusetts 02118. *Received February 14, 1973*. Presented in part at the 62nd Meeting of the Federation of American Societies for Experimental Biology (Schmid *et al.*, 1971a). Preliminary notes on the amino acid sequence of CNBr-I of α_1-acid glycoprotein were published (Schmid *et al.*, 1971b, 1972). This study was supported by grants from Hoffmann-La Roche, Inc., Nutley, N. J., the National Science Foundation (GB-24813), and the National Institute of General Medical Sciences (GM-10374), U. S. Public Health Service.

[1] Abbreviations used are: CNBr, cyanogen bromide; Gdn·HCl, guanidine hydrochloride; CM, carboxymethyl; BAW, *n*-butyl alcohol–acetic acid–water (200:30:75); BAWP, *n*-butyl alcohol–acetic acid–water–pyridine (15:3:10:12): BPW, *n*-butyl alcohol–pyridine–water (1:1:1); Glu, pyrrolidonecarboxylic acid or pyroglutamic acid residue.

ml per OD_{280} unit × 1000. Because of the large activity losses incurred during further purification, as judged by preliminary experiments, the enzyme was not further purified. The use of such a preparation was possible since the enzyme and the impurities, after the reaction with CNBr-I, could be separated readily from this substrate as described in the following.

Enzymatic Cleavage of Pyrrolidonecarboxylic Acid from CNBr-I. CNBr-I (100 mg) was dissolved in 5 ml of 0.05 M potassium phosphate buffer, pH 7.3, containing 0.01 M mercaptoethanol, mixed with 5 ml of the above-described enzyme solution, and incubated at 30°. After 8 hr, an additional 5 ml of the enzyme solution was added, and after 20 hr the reaction was terminated by adjusting the pH of the digest to 5.4 with acetic acid. For the isolation of the modified CNBr-I, the digest was applied to a Sephadex A-50 column (2.6 × 30 cm, 0.1 M sodium acetate buffer, pH 5.4) which retained the modified CNBr-I because of the sialic acid content (8%), while the enzyme and the impurities passed through the column. Using a NaCl gradient from 0.005 to 0.5 M in the mentioned buffer, CNBr-I was eluted subsequently, the effluent being monitored by the phenol–H_2SO_4 procedure (Spiro, 1966). Desialization of the modified CNBr-I was carried out as described above.

Reduction and S-Carboxymethylation. The reduction of the desialyzed α_1-acid glycoprotein (4.5 g) was carried out at pH 10 and 25° in the presence of 6 M Gdn·HCl, 0.2 M Tris-HCl buffer, and 3.0 M β-mercaptoethanol for 17 hr (Ishiguro *et al.*, 1969) and the subsequent carboxymethylation was carried out at pH 8.0 (Crestfield *et al.*, 1963). The excess reagents were removed from the modified protein by passage through an appropriate Sephadex G-25 column.

Chymotryptic Digestion. A solution of 4.5 g of reduced and alkylated desialyzed glycoprotein in 150 ml of water was adjusted to pH 8.0 with 1.0 N NaOH and mixed at 37° with 45 mg of chymotrypsin dissolved in 1.0 ml of water. The pH of the digest was kept constant with a pH-Stat (Radiometer, Copenhagen, Model TTT1c) using 1.0 N NaOH and N_2. After 60 min an additional 45 mg of the enzyme was added, and after 3 hr the reaction was terminated by lyophilization of the digest. The resulting residue dissolved in 50 ml of 10% acetic acid was passed through a Sephadex G-50 column (7.5 × 100 cm, 5% acetic acid) for the separation of the peptides from the glycopeptides (see Figure 1). These compounds are designated hereafter with the prefixes Ch and ChG, respectively.

Fractionation and Purification of the Carbohydrate-Free Chymotryptic Peptides. For this purpose chromatography on Dowex 50-X2 and Dowex 1-X2 (Schroeder, 1972), gel filtrations through Sephadexes G-10, -15, and -25, high-voltage electrophoresis at pH 6.5 in pyridine acetate buffer, and paper chromatography in BAW (*n*-butyl alcohol–acetic acid–water, 200:30:75), BAWP (*n*-butyl-alcohol–acetic acid–water–pyridine, 15:3:10:12), and/or BPW (*n*-butyl alcohol–pyridine–water, 1:1:1) were utilized. The chromatographic elution patterns were usually determined by measuring the ninhydrin color (570 nm) of aliquots of the effluents after alkaline hydrolysis (Yemm and Cocking, 1955). Appropriate fractions were pooled, concentrated under reduced pressure, and adjusted to a known volume of 5% acetic acid or water and stored frozen. The yields of the peptides were based on the amino acid content determined on 24-hr hydrolysates by a Technicon autoanalyzer.

Fractionation and Purification of the Chymotryptic Glycopeptides. For this purpose chromatography on DEAE- and CM-celluloses and Dowex 50 was used. The effluent was usually analyzed for neutral hexoses by the phenol–sulfuric acid procedure (Spiro, 1966). The separation and isolation of these glycopeptides were straightforward, and a concise description is found under Results.

Trifluoroacetylation of CNBr-I and Tryptic Digestion of CF_3CO-CNBr-I. This substitution reaction was carried out essentially according to Goldberger and Anfinsen (1962). For its enzymatic digestion CF_3CO-CNBr-I (6 g), which was grossly aggregated as judged by ultracentrifugal analysis, was suspended in 60 ml of 10 M recrystallized urea containing 0.1 M NH_4HCO_3, pH 8.6, and then heated for 10 min at 80° in order to completely dissolve the modified protein fragment. After cooling, 240 ml of 0.1 M NH_3HCO_4, pH 8.6, was added to decrease the urea concentration to 2 M so that subsequently the tryptic digestion (140 ml of 0.1% enzyme, 0.01 N HCl containing 0.01 M $CaCl_2$) could be carried out. Following incubation at 37° for 3 hr, a further 56 ml of this enzyme solution was added, and after 5 hr the digest was lyophilized. If CF_3CO-CNBr-I was not disaggregated in the presence of urea, the enzymatic digestion reached at most 70% of the expected value and 30% of the substrate was recovered as high molecular weight constituents on subsequent gel filtration.

Isolation of the Tryptic Peptides and Glycopeptides of CF_3CO-CNBr-I. The gross fractionation of the digest of CF_3CO-CNBr-I was achieved by gel filtration (Figure 4A) yielding three main fractions. Fraction 4A-1 (0.61 g) eluted at the void volume contained high molecular material. Fraction 4A-2 (4.20 g) contained the glycopeptides, designated with the prefix TG, and fraction 4A-3 contained the carbohydrate-free peptides, designated with the prefix T, which after desalting by passage through a G-10 column weighed 0.52 g. The total recovery was 5.33 g or 90%.

For removal of the trifluoroacetyl residues, fraction 4A-2 was dissolved in 84 ml of 1 M piperidine and allowed to stand at 0° for 2 hr. Thereafter, the pH of the solution was adjusted to 8.6 using 50% acetic acid followed by desalting by gel filtration through a Sephadex G-10 column (3 × 75 cm, 0.1 M NH_4HCO_3, pH 8.6) and lyophilization (yield, 4.1 g or 97%). Fraction 4A-3 was treated by the same procedure.

Desialyzation of the detrifluoroacetylated glycopeptide fraction was carried out as indicated above. The resulting preparation was dissolved in 30 ml of 0.1 M NH_4HCO_3, pH 8.6, and desalted on a Sephadex column (Figure 4B) yielding three fractions. The minor fraction eluted at the void volume contained high molecular weight material and was not further studied. Fraction 4B-1, which included the large glycopeptides, weighed 0.25 g and fraction 4B-2 weighed 3.28 g. The cleaved sialic acid (*N*-acetylneuraminic acid) and salts (0.77 g) were eluted last. For the isolation of the peptides and glycopeptides the techniques used for the fractionation of the chymotryptic peptides were employed.

Peptic Peptides. A series of peptides (prefix P) and glycopeptides (prefix PG) was isolated in homogeneous form from a peptic digest of desialyzed α_1-acid glycoprotein employing the procedures mentioned above. This enzymatic hydrolysis was carried out at pH 2.0, 37°, for 7 hr at an enzyme–substrate ratio of 1:50 and a protein concentration of 3%.

Homogeneity. This property of the peptides was established by high voltage electrophoresis at two pH values (6.5, 1.8, and/or 9.5) and by paper chromatography in BAW (3:1:1) or BAWP and was also evaluated from their amino acid compositions. All electrophoretograms and paper chromatograms were stained first with ninhydrin, then with the

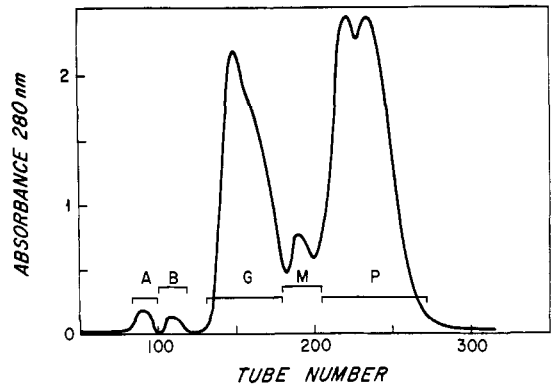


FIGURE 1: Gel filtration on Sephadex G-50 of a chymotryptic digest of desialized, reduced, and carboxymethylated α_1-acid glycoprotein. The column (7.5 × 97 cm) was equilibrated with 5% acetic acid. The flow rate was adjusted to 80 ml/hr. The effluent which was passed through a Uvicord to record the absorption at 280 nm was collected in 14-ml fractions.

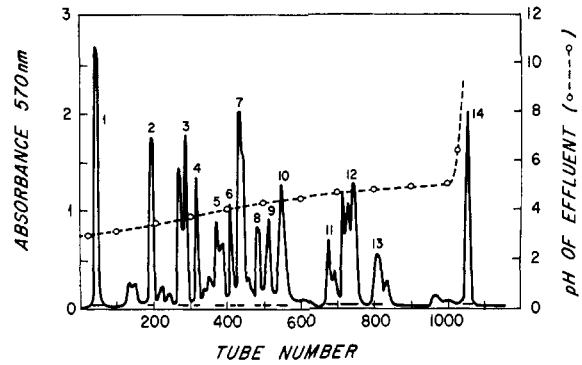


FIGURE 2: Chromatography of the peptide fraction P on Dowex 50. The peptide mixture (1.3 g), fraction P (Figure 1), was dissolved in 45 ml of 0.2 M pyridine–acetate buffer, pH 3.1, the pH of the mixture was lowered to 2 with a few drops of HCl, and the mixture was applied to a column (3 × 220 cm) of Dowex 50-X2 (200–400 mesh) equilibrated with 0.2 M pyridine–acetate, pH 3.1. The acidic peptides were eluted with 500 ml of the same buffer, then a linear gradient was applied consisting of 7 l. of 0.2 M pyridine–acetate buffer, pH 3.1, and 9 l. of 2 M pyridine–acetate buffer, pH 5.3. After the pH of the effluent reached a value of 5.1, the gradient buffer was replaced by 4 N NH_4OH which yielded one further peptide. Fractions of 16 ml were collected at a flow rate of 180 ml.

Sakaguchi and/or Ehrlich reagent, followed by the starch iodine procedure (Ikenaka *et al.*, 1972). As references for the high voltage electrophoresis at pH 6.5, aspartic acid and lysine were used so that the electrostatic net charges of the peptides could be assessed. It is important to note that the apparent electrophoretic mobilities of the tryptic glycopeptides were always low and similar to each other, although their amino acid compositions may be quite different. Therefore, the above criteria of purity were not considered to be sufficient to demonstrate true homogeneity of the latter glycopeptides. Hence, the homogeneity of these compounds was evaluated primarily from analyses of their amino- and carboxyl-terminal amino acids.

Analytical Techniques. These procedures including the methods for determining the amino acid sequences were described in the previous paper of this study (Ikenaka *et al.*, 1972). The assumptions that all tryptic peptides containing arginine would have this residue as their carboxyl terminus and that the arginine-free tryptic peptides with lysine possess this amino acid as their carboxyl terminus were proven by carboxyl-terminal amino acid analyses. For the determination of the carboxyl-terminal amino acids by the action of carboxypeptidase A on the glycopeptides (when available only in small quantities), it proved to be useful to separate initially the liberated amino acids from the modified glycopeptides by gel filtration through Sephadex G-25 (1 × 40 cm) followed by paper chromatography in appropriate solvent systems.

Results

Preparation of the Chymotryptic Peptides and Glycopeptides

Separation of the Peptides from the Glycopeptides. The chymotryptic digest (4.5 g) described above yielded on gel filtration (Figure 1) three major fractions, designated G (2.30 g of glycopeptides), M (0.43 g of a mixture of glycopeptides and large peptides), and P (1.34 g of peptides). Small amounts of apparently undigested (fraction A, 29 mg) and partially digested (fraction B, 17 mg) protein eluted first were not further studied. Rechromatography of fraction M on the same Sephadex column afforded a glycopeptide fraction which was combined with fraction G and a peptide fraction which contained the relatively large carbohydrate-free peptides.

Isolation of the Carbohydrate-Free Peptides. Chromatography of fraction P on a Dowex 50 column afforded 14 subfractions (Figure 2) which were further chromatographed on Dowex 1-X2 or Sephadex G-10 or -25 (a representative pattern is shown in Figure 3) in order to obtain the major peptides in homogeneous form. For the purification of the mentioned large peptides the same procedures were utilized. The amino acid compositions and recoveries of these peptides are listed in Table I.

Isolation of the Glycopeptides. The glycopeptide mixture (fraction G, Figure 1, 900 mg) was chromatographed on a DEAE-cellulose column (2.5 × 55 cm) yielding three main fractions whose recoveries accounted for 75%. These fractions appeared heterogeneous as judged by high-voltage electrophoresis and amino acid analyses. Further purification was achieved by chromatography on CM-cellulose

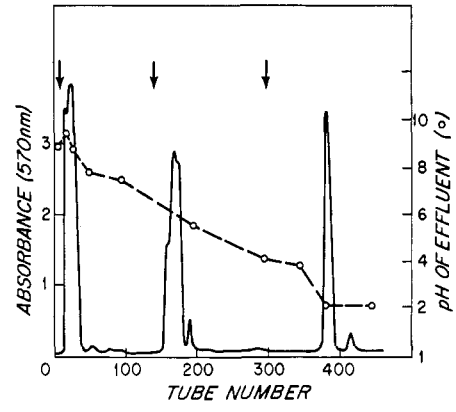


FIGURE 3: Ion exchange chromatography of fraction 3 on Dowex 1-X2 (1 × 100 cm) using *N*-ethylenemorpholine–pyridine–H_2O–AcOH buffer, pH 9.4. This peptide fraction was applied in 2 ml of buffer. After elution with 60 ml of the same buffer, three successive gradients were applied at the positions indicated by the arrows. These gradients were made with the original buffer and 0.1 N, 0.5 N, and 25% acetic acid, respectively, as indicated in the figure. The flow rate was 18 ml/hr and the effluent was collected in 3-ml fractions.

TABLE I: Amino Acid Composition of Carbohydrate-Free Chymotryptic Peptides Derived from α_1-Acid Glycoprotein.

Amino Acid	Ch-1	Ch-2	Ch-3	Ch-4	Ch-5	Ch-6	Ch-7	Ch-8	Ch-9	Ch-10	Ch-11	Ch-12
CM-Cysteine					0.81 (1)							
Aspartic acid		1.01 (1)		0.94 (1)	0.98 (1)					0.99 (1)	0.95 (1)	1.00 (1)
Threonine		0.94 (1)			0.86 (1)					1.64 (2)	1.80 (2)	1.72 (2)
Serine			0.81 (1)									
Glutamic acid	0.97 (1)			0.95 (1)	2.84 (3)	1.98 (2)			2.10 (2)			
Proline	0.93 (1)											
Glycine		0.93 (1)				2.00 (2)			2.01 (2)			
Alanine			1.88 (2)				0.97 (1)	1.00 (1)	1.03 (1)			
Valine						0.98 (1)			1.16 (1)			
Isoleucine	1.00 (1)	1.07 (1)	1.00 (1)	1.00 (1)	0.92 (1)					0.89 (1)		
Leucine	1.04 (1)						1.01 (1)	1.80 (2)	1.80 (2)	1.00 (1)	1.95 (2)	
Tyrosine					0.94 (1)					0.85 (1)		
Phenylalanine			0.91 (1)			1.00 (1)						0.93 (1)
Lysine		1.10 (1)								0.84 (1)	1.05 (1)	1.10 (1)
Histidine							0.95 (1)	0.84 (1)	2.04 (2)	0.95 (1)		
Arginine		1.20 (1)		1.00 (1)	1.07 (1)	1.02 (1)			1.20 (1)		1.10 (1)	1.00 (1)
Tryptophan		0.98 (1)										
Total no. of residues	4	7	5	4	9	7	3	4	11	8	7	6
Yield (%)	80	5	3	45	41	38	3	24	9	6	17	24
Position in sequence	1–4	19–25	28–32	62–65	66–74	92–98	99–101	99–102	92–102	103–110	104–110	105–110

TABLE II: Amino Acid Composition of Tryptic Carbohydrate-Free Peptides of CNBr-I of α_1-Acid Glycoprotein.

Amino Acid	T-1	T-2	T-3	T-4	T-5	T-6	T-7	T-8	T-9	T-10	Variant T-10
CM-Cysteine											
Aspartic acid									1.13 (1)	0.73 (1)	1.00 (1)
Threonine		0.96 (1)	0.83 (1)	0.90 (1)					0.90 (1)	1.72 (2)	1.70 (2)
Serine		0.97 (1)									
Glutamic acid	0.89 (1)			2.21 (2)	1.20 (1)	2.12 (2)	1.95 (2)	1.00 (1)			
Proline	0.92 (1)										
Glycine		1.00 (1)			2.14 (2)	1.95 (2)	2.21 (2)				

Alanine		2.14 (2)	2.10 (2)			1.15 (1)		1.00 (1)					
Valine						1.04 (1)	1.10 (1)	0.40 (0.4)					
Isoleucine	1.00 (1)	2.12 (2)	0.99 (1)			1.03 (1)		0.63 (1)					
Leucine	1.25 (1)					3.10 (3)		3.00 (3)					0.90 (1)
Tyrosine		1.25 (1)	0.91 (1)	1.00 (1)	0.87 (1)	0.83 (1)	0.91 (1)			0.93 (1)			
Phenylalanine		1.98 (2)	1.72 (2)			1.12 (1)	1.01 (1)	1.43 (1.5)					
Lysine		1.13 (1)						0.28 (0)	1.00 (1)	1.00 (1)	1.10 (1)		
Histidine						1.96 (2)	1.92 (2)	1.82 (2)					
Arginine		1.10 (1)	1.00 (1)	1.00 (1)	1.00 (1)	1.00 (1)		1.14 (1)					
Tryptophan[a]			*b* (1)	0.89[a] (1)									
Homoserine										0.70 (1)	0.70 (1)		
Total no. of residues	4	13	9	5	5	15	10	10	3	6	6		
Yield (%)	6	2	46	15	9	52	5	2	3	16	1		
Position in sequence	1–4	21–33	25–33	64–68	91–95	91–105	91–110	96–105	106–108	106–111	106–111		

[a] Calculated from 280-nm absorbance. [b] Ehrlich-positive. Amino acids whose content amounted to 0.1 mol or less are not listed.

columns at pH 4.0 of the two relatively acidic glycopeptide fractions and at pH 5.0 of the third glycopeptide fraction. For these separations ionic strength gradients from 0.001 to 2.0 were employed. In each case a heterogeneous fraction was obtained first followed by a homogeneous glycopeptide. From these fractions the remaining glycopeptides were isolated by chromatography on CM-cellulose and Dowex 50. The seven glycopeptides which were isolated in homogeneous form were analyzed for their amino acid compositions (Table III) and amino-terminal amino acids.

Preparation of the Tryptic Peptides and Glycopeptides

The separation of the carbohydrate-free, detrifluoroacetylated tryptic peptides yielded 11 pure compounds (Table II). One of these peptides (T-3) was insoluble under the conventional conditions but could be chromatographed on Sephadex G-100 in 30% acetic acid.

The large tryptic glycopeptides (fraction 4B-1, Figure 4B) were separated from each other by chromatography on a CM-cellulose column (200 mg, 1.8 × 60 cm, pH 5.0, 0.005 M pyridine acetate, ionic strength gradient to 0.2 M) yielding two fractions containing TG-4 and TG-8, respectively. The glycopeptides of fraction 4B-2 (Figure 4B) were purified by chromatography on a GE-cellulose column (Figure 4C) affording a major fraction (4C-1, 2.7 g) and two minor fractions (4C-2 and 4C-3, 0.26 and 0.31 g, respectively). On high-voltage electrophoresis, the latter two fractions appeared to be considerably more heterogeneous than fraction 4C-1 and were, therefore, not investigated further. Fractionation of the glycopeptides of fraction 4C-1 was achieved first by chromatography on an SE-cellulose column (Figure 4D) yielding six fractions whose weights were 128, 503, 101, 127, 980, and 431 mg, respectively. Each subfraction was rechromatographed on an SE-cellulose column using appropriate pH–ionic strength gradients. The resulting main fractions were further chromatographed on Dowex 1-X2 (*e.g.*, Figure 4E) according to Schroeder (1972) excepting that of fraction D6 which was subjected to CM-cellulose chromatography (Figure 4F) (for amino acid composition see Table IV). It is of interest that the glycopeptides TG-5, TG-6, and TG-7 which differed only in their phenylalanine contents could be separated from each other to a surprisingly high degree.

Amino Acid Sequence of CNBr-I

In the following the elucidation of the amino acid sequences of the chymotryptic peptides and glycopeptides is described in detail (see also Figure 5). However, the sequences of the tryptic[2] and peptic peptides and glycopeptides are presented only in Figure 5 which shows the overlaps with the former series of peptides. A few comments pertinent to the study of these tryptic compounds are found in the Discussion.

The positions of the carbohydrate units were identified as follows (for further details see below). For the four glycopeptides, ChG-4, ChG-5, ChG-6, and ChG-7, the direct Edman procedure was employed. All phenylthiohydantoins, including threonine, serine, and lysine that are located in the proximity of the carbohydrate–peptide linkage, revealed the properties expected for these amino acid residues, without

[2] As indicated in Figure 5 the ϵ-amino groups of the lysine residues were trifluoroacetylated to varying degrees. The very incomplete substitution of residue 39 and the almost complete reaction of residue 55 can probably not be explained by steric hindrance by the carbohydrate units, because both these residues are located adjacent to a carbohydrate–asparaginyl bond.

TABLE III: Amino Acid Composition of Chymotryptic Glycopeptides Derived from α_1-Acid Glycoprotein.

Amino Acid	ChG-1	ChG-2	ChG-2a Arg-, variant	ChG-3	ChG-4	ChG-6	ChG-7
CM-Cysteine	0.80 (1)	0.79 (1)	0.70 (1)				
Aspartic acid	2.14 (2)	3.19 (3)	3.17 (3)	2.05 (2)	1.95 (2)	1.00 (1)	2.03 (2)
Threonine	2.02 (2)	3.05 (3)	2.84 (3)	0.92 (1)	2.76 (3)	1.30[b] (1.3)	1.03 (1)
Serine				0.93 (1)		0.61[b] (0.6)	0.96 (1)
Glutamic acid		1.00 (1)		4.56 (2)	1.01 (1)		1.91 (2)
Proline	2.17 (2)	1.73 (2)	1.77 (2)		1.00 (1)		
Glycine		1.05 (1)	1.02 (1)				1.02 (1)
Alanine	1.96 (2)	2.09 (2)	1.95 (2)	0.97 (1)			
Valine	1.82 (2)	1.98 (2)	2.01 (2)	0.95 (1)			1.26[b] (1.3)
Isoleucine	1.00 (1)	1.98 (2)	1.83 (2)	0.94 (1)	0.90 (1)		0.73[b] (0.7)
Leucine	1.92 (2)	2.04 (2)	1.92 (2)			0.30[b] (1)	0.49[b] (0.5)
Tyrosine				0.91 (1)	0.29[a]	0.85 (1)	1.00 (1)
Phenylalanine		0.25 (1)	0.21 (1)	1.00 (1)	1.77 (2)		
Lysine		1.09 (1)	1.00 (1)	1.02 (1)	1.02 (1)		
Arginine			0.88 (1)	0.94 (1)			1.89 (2)
Tryptophan		ND (1)	ND (1)				
Total no. of residues	14	22	22	16	12	5	13
Yield (%)	25	18	5	34	34	27	62
Position in sequence	5–18	5–26	5–26	33–48	50–61	75–79	79–91

[a] Glycopeptide Ch-G4 may represent a mixture of two peptides, one of which is lacking the amino-terminal tyrosine residue. [b] See the text for an explanation.

TABLE IV: Amino Acid Composition of Tryptic Glycopeptides Derived from CNBr-I of α_1-Acid Glyocprotein.

Amino Acid	TG-3[a]	TG-4	TG-5	TG-6	TG-7	TG-8	TG-9	Variant TG-9	TG-10
CM-Cysteine							0.75 (1)	0.65 (1)	
Aspartic acid	1.82 (2)	2.25 (2)	2.05 (2)	2.04 (2)	2.00 (2)	4.00 (4)	3.20 (3)	3.33 (3)	1.32 (1)
Threonine		3.70 (4)	2.81 (3)	2.93 (3)	3.20 (3)	3.90 (4)	1.83 (2)	0.5[b]	1.10 (1)
Serine		1.10 (1)				0.93 (1)		1.4[b] (2)	0.85 (1)
Glutamic acid	1.92 (2)	3.68 (4)	1.20 (1)	1.20 (1)	0.94 (1)	5.74 (6)	3.40 (3)	3.10 (3)	0.88 (1)
Proline		1.28 (1)	1.02 (1)	1.18 (1)	1.00 (1)	1.00 (1)			
Glycine									1.05 (1)
Alanine		1.29 (1)				0.93 (1)			
Valine		1.10 (1)				0.73 (1)	1.00 (1)	1.20 (1)	0.37[c]
Isoleucine		1.93 (2)	1.20 (1)	1.10 (1)	1.19 (1)	1.65 (2)	1.00 (1)	0.45[b]	0.80 (1)[c]
Leucine		1.33 (1)	1.20 (1)	1.19 (1)	1.18 (1)	1.00 (1)	1.00 (1)	1.25 (1)	
Tyrosine	0.82 (1)	1.10 (1)	0.92 (1)	0.93 (1)	1.00 (1)	2.13 (2)	1.93 (2)	2.03 (2)	
Phenylalanine		3.65 (4)	2.20 (2)	2.92 (3)	3.73 (4)	4.20 (4)		1.00 (1)[b]	
Lysine	1.00 (1)	1.20 (1)	1.00 (1)	1.00 (1)	0.78 (1)	2.01 (2)			
Histidine									
Arginine		1.00 (1)	0.95 (1)	0.89 (1)	1.00 (1)	1.00 (1)	1.00 (1)	1.00 (1)	1.00 (1)
Tryptophan									
Total no. of residues	6	24	14	15	16	30	15	15	7
Yield (%)	43	4	19	20	12	1	40	13	50
Position in sequence	34–39	40–63	50–63	49–63	48–63	34–63	69–83	69–83	84–90

[a] The fractions containing the glyocpeptides TG-1 and TG-2 are discussed in the text. [b] For variants see the text. [c] A variant peptide with valine in place of isoleucine is discussed in the text.

substitution of their reactive groups, such as migration rates on thin-layer chromatography, absorption spectra, and apparent electrophoretic mobilities on high-voltage electrophoresis. The only exceptions were the asparagine residues. These residues, because they each carried a carbohydrate unit, remained in the final aqueous solution. For other glycopeptides, Th-G of ChG-1 and ChG-3, the subtractive Edman technique was applied. The peptides obtained after each degradation step were analyzed for neutral hexose. The obtained results together with enzymatic degradation of the

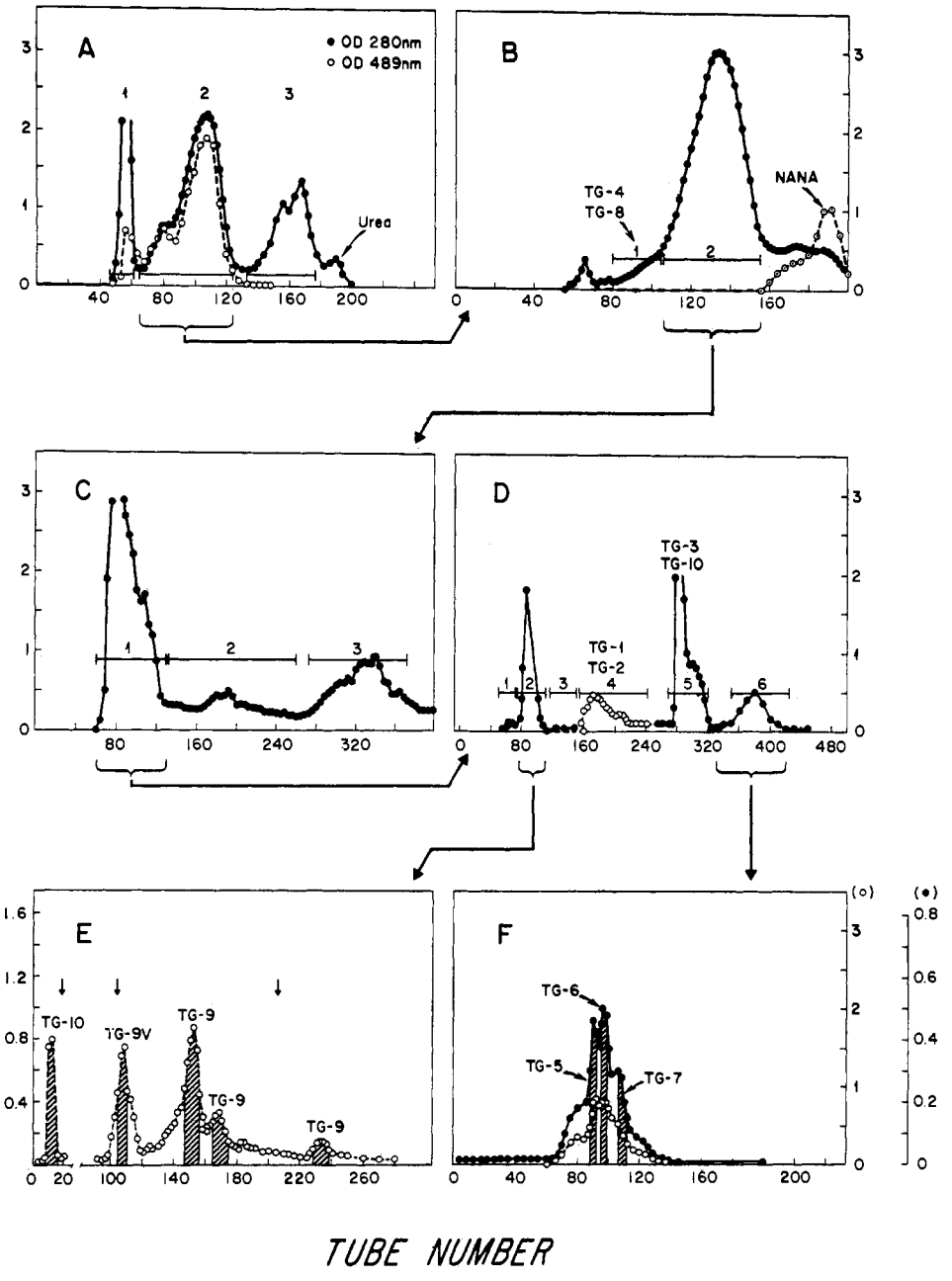


FIGURE 4: Gel filtration of a tryptic digest of CF_3CO-CNBr-I of α_1-acid glycoprotein through a Sephadex G-50 column (4.6 × 1.45 cm, 3 g dissolved in 90 ml of 0.1 M Na_4HCO_3, pH 8.6; flow rate 30 ml/hr; 10 ml per tube). (A) Note that after fraction 4A-3, urea was eluted at the 190th tube. In all figures, the tube number is indicated on the abscissa, while the absorbance at 280 nm (●), the relative neutral hexose content (○), and/or the peptide content (570 nm) are given on the ordinate. (B) Gel filtration of the tryptic glycopeptides of CNBr-I α_1-acid glycoprotein, after removal of the trifluoroacetyl residues and desialyzation through a Sephadex G-50 column (4.6 × 145 cm) in 0.01 M NH_4HCO_3, pH 9.0. The column was washed with 1800 ml of the same solution; ionic strength gradient of constant pH, using 1.5 l. of 0.01 M and 1.5 l. of 0.3 M NH_4HCO_3 solution (pH 9.0). The free sialic acid (⊙) was monitored by the resorcinol procedure of Svennerholm (1963). (C) Chromatography of fraction 4B-2 (Figure 1B) on a GE-cellulose column (2.5 × 100 cm; 0.01 M NH_4HCO_3, pH 9.0; flow rate 33 ml/hr; GE-cellulose was Bio-Rad No. 4396, 0.24 mequiv/g). (D) Chromatography of the low molecular glycopeptides (fraction 4C-1, Figure 1C) on a SE-cellulose column (1 × 125 cm, 0.05 M formic acid, Bio-Rad No. 5832). After washing the column with 1.7 l. of 0.05 M HCOOH, an ionic strength pH-gradient prepared by using 1.5 l. of 0.05 M HCOOH and 1.5 l. of 0.3 M pyridine acetate, pH 3.7, was applied. The second gradient was applied after the third peak was eluted and was made as follows: 1.0 l. of 0.3 M pyridine acetate, pH 3.7, and 1.0 l. of 3 M pyridine acetate, pH 3.7. (E) Chromatography of certain tryptic glycopeptides (fraction 4D-2, Figure 1D) on a Dowex 1-X2, column 1.2 × 80 cm. The peptide mixture (100 mg) was applied in pH 9.4 ethylenemorpholine–pyridine–acetic acid–water buffer. At the positions indicated by the first two arrows the pH value of this buffer was changed to 8.4 and 6.4, respectively. At the position indicated by the third arrow 1 N acetic acid was applied. (F) Chromatography of glycopeptide fraction 4D-6 (Figure 4D) on a CM-cellulose column (for further details see text).

carboxyl terminus of ChG-1 allowed us to establish unambiguously the asparaginyl residue to which the carbohydrate unit is linked. Two further short glycopeptides (ThG-NG of ChG-3 and TG-3) were degraded with enzymes (carboxypeptidase A and B and Pronase, respectively) to such an extent that only an asparagine residue remained with the carbohydrate. After the enzymatic degradation, the digest was separated by Sephadex G-25 filtration into a glycopeptide and an amino acid fraction which were then analyzed for amino acids. In the subsequent presentation the asparaginyl

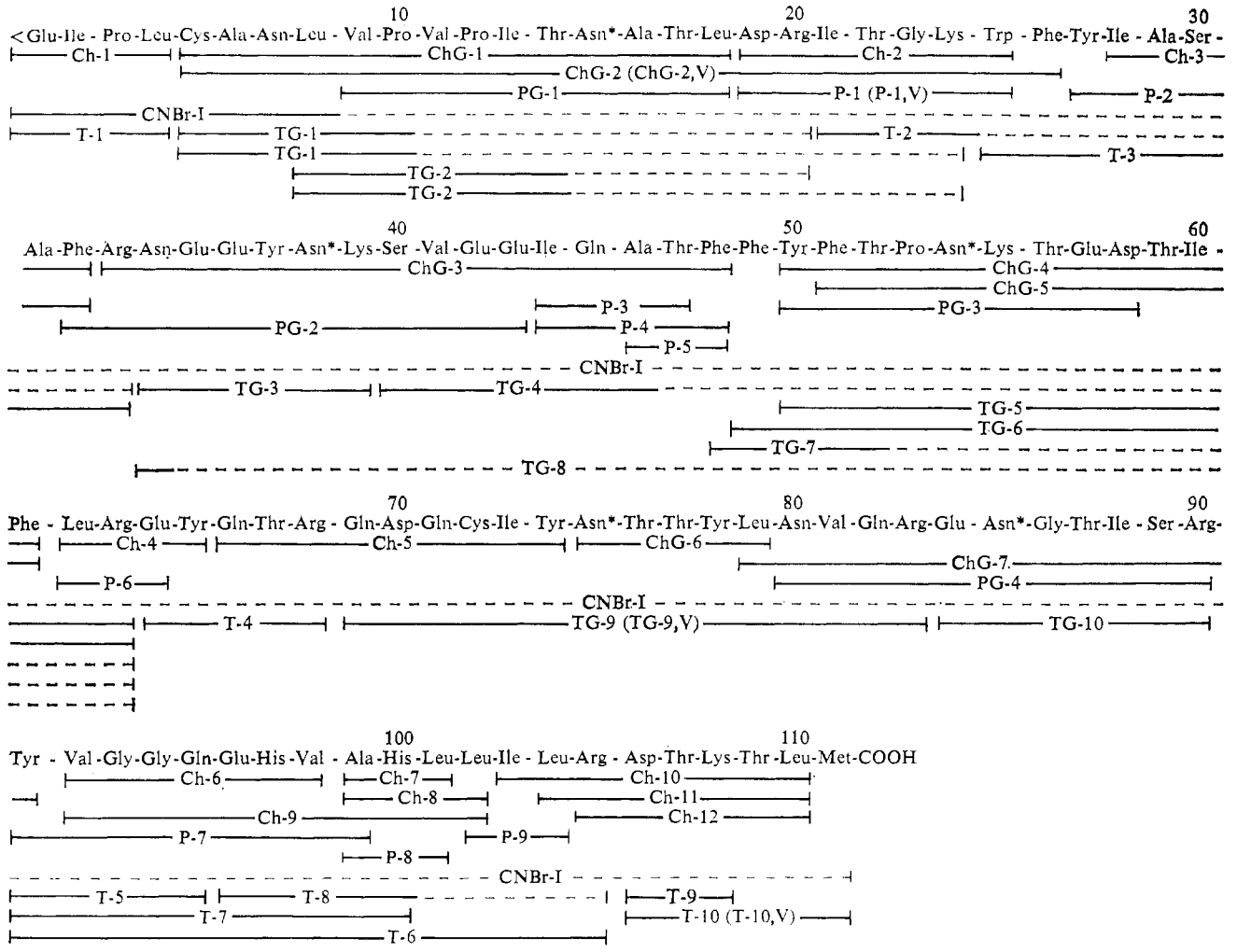


FIGURE 5: The amino acid sequence of the amino-terminal CNBr fragment of α_1-acid glycoprotein and the sequences of the chymotryptic, tryptic, and peptic peptides and glycopeptides on which the structure of this fragment is based. Abbreviations used are: Ch, chymotryptic; P, peptic; T, tryptic; and G, glycopeptides. The peptides whose symbols appear in parentheses are those that contain the bottom one of the two variants at any variable position in the sequence. If the variant peptides could not be separated from each other, their symbols are not in parentheses.

residues that carry a carbohydrate unit are marked with an asterisk (Asn*). These results are in agreement with the earlier report by Anderson *et al.* (1964).

Partial Sequence of CNBr-I (Residues 1–8). On incubation with the highly specific enzyme, pyrrolidonecarboxylylpeptidase, approximately 50% of the amino-terminal residue of CNBr-I, pyrrolidonecarboxylic acid was cleaved. This finding represents direct evidence that pyrrolidonecarboxylic acid and not an acylated amino acid forms the amino terminus of this fragment and, hence, of α_1-acid glycoprotein and thus confirms our earlier report (Ikenaka *et al.*, 1966) and was recently confirmed by Szewczuk and Kwiatkowska (1970). Together with the initial enzymatic degradation, the sequence of eight amino-terminal residues could be established

<Glu-Ile-Pro-Leu-Cys-Ala-Asn-Leu

0.50[3] 0.50 0.42 0.34 0.27 0.20 0.15 0.10

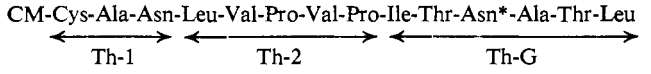
This sequence and those described below are compiled in Figure 5.

Peptide Ch-1 (Residues 1–4). This ninhydrin-negative peptide isolated from fraction 1 (Figure 2) and purified by passage through a Sephadex G-25 column was treated with 1 N NaOH at 25° for 80 hr in order to transform its pyrrolidonecarboxylic acid to a glutamyl residue (Ikenaka *et al.*, 1966). After neutralization and desalting by gel filtration through an appropriate Sephadex column, the sequence of the modified peptide was established by three subtractive Edman steps (Elzinga *et al.*, 1965). Particular care was taken to identify isoleucine and leucine. From these data it was concluded that the sequence of the original peptide was <Glu-Ile-Pro-Leu, representing the amino terminus of α_1-acid glycoprotein.

Glycopeptide ChG-1 (Residues 5–18). Seven steps of the subtractive Edman and carboxypeptidase A digestion of this glycopeptide established both the amino- (CM-Cys-Ala-Asx-Leu-Val-Pro-Val-) and carboxyl- (-Thr-Leu) terminal sequences. The amino acid composition of the carboxypeptidase A degraded glycopeptide was found to be in agreement with the established sequence presented below. Digestion with thermolysis of ChG-1, followed by gel filtration through a Sephadex G-15 column (1 × 100 cm), yielded a glycopeptide and a fraction containing two peptides which were resolved by paper chromatography in BAW. The sequence of peptide

[3] The number under each amino acid residue indicates the molar recovery of the residue.

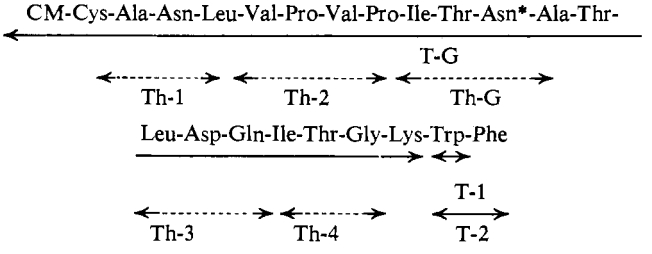
Th-1 which consisted of equimolar amounts of CM-cysteine, alanine, and aspartic acid was established by the subtractive Edman technique. Analysis (Blackburn, 1968) of a complete aminopeptidase M digest of Th-1 demonstrated the presence of asparagine. Peptide Th-2, which was composed of valine (1.44 (2)), leucine (0.56 (1)), and proline (2.00 (2)), revealed on subtractive Edman degradation the sequence indicated below and represented an internal section of ChG-1. The sequence of these thermolysin peptides within ChG-1 was confirmed by the above-described amino-terminal sequence of the original glycopeptide. The resulting glycopeptide (Th-G) which represents the carboxyl terminus of ChG-1 was elucidated by four steps of the subtractive Edman, the amino acid composition, and the above-mentioned carboxyl terminus of ChG-1. Hence, the sequence of ChG-1 is



Peptide Ch-2 (*Residues 19–25*). This Ehrlich-positive peptide isolated from fraction 14 (Figure 2) yielded on digestion with carboxypeptidase A (4 hr) only tryptophan. A total digest with aminopeptidase M (20 hr) did not afford any asparagine. Six steps of the subtractive Edman thus established the sequence of this peptide as follows: Asp-Arg-Ile-Thr-Gly-Lys-Trp.

Glycopeptide ChG-2 (*Residues 5–26*). This glycopeptide possessed an amino acid composition similar to that of ChG-1 except for eight additional residues. Since it contained only one residue of lysine, ChG-2 was first digested with trypsin (5.3 mg in 1 ml of 0.05% NH_4COH_3, pH 8.0, 50 μg of enzyme, 37°, 20 hr) yielding on gel filtration (Sephadex G-25, 1 × 100 cm, 1% acetic acid) a glycopeptide (T-G) and a mixture of two compounds (T-1 being Trp and T-2 being Trp-Phe). Using carboxypeptidase B it could be shown that T-G possessed carboxyl-terminal lysine, and employing the dansyl–Edman procedure (Gray, 1967) to contain amino-terminal CM-cysteine. Further, stepwise degradation by the direct Edman procedure revealed the amino terminus to be CM-Cys-Ala-Asn-.

The glycopeptide (T-G) was next subjected to cleavage by thermolysin and afforded on gel filtration a glycopeptide (ThG) and a mixture of four peptides which was fractionated by paper chromatography. The amino acid compositions of these four compounds revealed that Th-1 was the amino-terminal peptide and demonstrated the position of Th-4 to be within ChG-2. With the aid of the sequence of the peptic peptide (P-1, see below), the position of Th-3 could also be established. The sequence of Th-4 was elucidated by the subtractive Edman procedure and found to be Ile-Thr-Gly-Lys. Three steps of the subtractive Edman of Th-G (Ile-Thr-Asx) established the position of this glycopeptide. Hence, the sequence of ChG-2 is

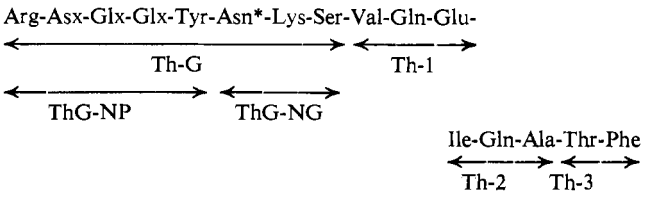


Glycopeptide ChG-2, Variant (*Residues 5–26*). This compound had the same amino acid composition as ChG-2 except for an additional *arginine* residue and the lack of *glutamine*. In view of the earlier report on the replacement of an arginine for a glutamine in this peptide sequence (Nimberg *et al.*, 1971) and in view of the corresponding peptic peptides described below, this glycopeptide was considered a variant of ChG-2.

Peptide Ch-3 (*Residues 28–32*). Fraction 3 (Figure 2) yielded on chromatography on Dowex 1 three subfractions, one of which contained a homogeneous peptide (Ch-3). The sequence of this Sakaguchi-negative, neutral peptide was elucidated with the use of the subtractive Edman technique, carboxypeptidase A, and aminopeptidase M. The resulting sequence, Ile-Ala-Ser-Ala-Phe, was confirmed by Dr. R. A. Laursen using the solid phase peptide sequencer (Laursen, 1971). For a *variant peptide* see the peptic peptides below. It should also be noted that residue 27 was not determined in this series of analysis, but was identified with the aid of the peptic peptides (see below).

Glycopeptide ChG-3 (*Residues 33–48*). The release of the amino acids from this Sakaguchi-positive glycopeptide by carboxypeptidase A measured as a function of the incubation time (0.5, 1.5, 2.5, 5.5, and 24 hr) revealed -Ala-Thr-Phe to be the carboxyl terminus. For these analyses the digests were separated by paper chromatography on BAW whereby the degraded glycopeptides remained immobile. The glycopeptide of the 24-hr digest, isolated by gel filtration (Sephadex G-25, 1 × 85 cm), was subsequently subjected to six steps of the subtractive Edman establishing the amino-terminal sequence: Arg-Asx-Glx-Glx-Try-Asn-*. At the sixth step the carbohydrate moiety was removed.

Thermolysin digestion of ChG-3 (4.2 mg) for 20 hr yielded on gel filtration (Sephadex G-15, 1 × 100 cm) a glycopeptide (ThG) and a peptide mixture which could be resolved on paper chromatography in BAW into three peptides. The amino acid compositions of these compounds were determined, and, using the subtractive Edman procedure, their sequences were established to be: Val-Gln-Glu (Th-1), Ile-Gln-Ala (Th-2), and the carboxyl-terminal dipeptide (Th-3) Thr-Phe. The glycopeptide (ThG) was next digested with Nagarse and the digest subjected to paper chromatography in BAW. The resulting glycopeptide (ThG-NG) which contained equimolar amounts of aspartic acid, serine, and lysine possessed carboxyl-terminal serine released by carboxypeptidase A and penultimate lysine subsequently released with carboxypeptidase B. The formed peptide (ThG-NP) consisted of aspartic acid (1), glutamic acid (2), tyrosine (1), and arginine (1), the latter being amino terminal (aminopeptidase M). Hence, the sequence of ChG-3 is



Glycopeptidase ChG-4 (*Residues 50–61*). Degradation by the direct Edman procedure afforded the sequence of ten amino acids: Tyr-Phe-Thr-Pro-Asn*-Lys-Thr-Glu-Asp-Thr. In view of the specificity of chymotrypsin and the amino composition of ChG-4, the carboxyl terminus of this peptide was assumed to be phenylalanine. The sequence of ChG-4 is therefore: Tyr-Phe-Thr-Pro-Asn*-Lys-Thr-Glu-Asp-Thr-Ile-

Phe. It should be noted that residue 49 is not included in this peptide or ChG-3, but its position was established in the study of the corresponding tryptic peptides.

Glycopeptide ChG-5 (Residues 51–61). This glycopeptide, whose amino acid composition is not listed in Table III, had the same sequence as ChG-4 except that it was lacking the amino-terminal tyrosine of ChG-4. Degradation by the direct Edman procedure and by carboxypeptidase A yielded the following partial sequence: Phe-Thr-Pro-Asn*-Lys-(Thr, Glu,Asp,Thr)-Ile-Phe. The formation of two very similar peptides which differed by one residue such as ChG-4 *vs.* ChG-5 has also been reported by Jackson and Hirs (1970) in their study on porcine ribonuclease and is expected from the specificity of this enzyme.

Peptide Ch-4 (Residues 62–65). This neutral peptide isolated from fraction 12 (Figure 2) was Sakaguchi-positive. The absence of glutamine was demonstrated by the analysis of an aminopeptidase M digest of the peptide. This finding and three steps of the subtractive Edman procedure established the sequence of Ch-4 as Leu-Arg-Glu-Tyr.

Peptide Ch-5 (Residues 66–74). Since the direct Edman degradation failed, it was assumed that this peptide isolated from fraction 2 (Figure 2) possessed an amino-terminal pyrrolidonecarboxylic acid residue. A short period of digestion with carboxypeptidase A afforded tyrosine (1.00) and isoleucine (0.20). Subsequently the peptide was digested with trypsin, lyophilized, and subjected to the direct Edman procedure yielding the following additional, partial sequence: Gln-Asp-Gln-CM-Cys. Further, a Sakaguchi-positive, ninhydrin-negative peptide which was isolated from this tryptic digest using high-voltage electrophoresis at pH 3.6 was composed of equimolar amounts of glutamic acid, threonine, and arginine. Again the Edman technique was not successful. Based on this information the sequence of Ch-5 is established unambiguously: Gln-Thr-Arg-Gln-Asp-Gln-Cys-Ile-Tyr.

Glycopeptide ChG-6 (Residues 75–79). Four steps of the direct Edman technique yielded the following sequence

$$\text{Asn*-Thr-}\underset{\text{Ser}}{\text{Thr}}\text{-Tyr-Leu}$$

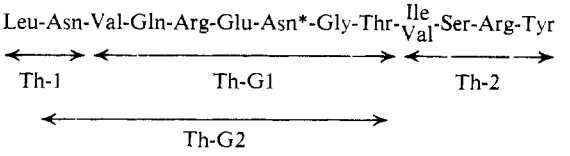
The sequence of the two tryptic glycopeptides, TG-9 and TG-9,-variant, confirmed this *amino acid substitution* and revealed further the same replacement in position 77 (Figure 5, Table VI).

Glycopeptide ChG-7 (Residues 79–91). Due to the particular carboxyl-terminal sequence of ChG-6 and the specificity of chymotrypsin, two very similar glycopeptides were present in this fraction. One of these peptides possessed an amino-terminal leucine, while the other had an asparaginyl residue, so that on degradation by the direct Edman method at each step an "overlapping" pair of amino acids was obtained. The partial amino-terminal sequence proved to be Leu-Asn-Val-Gln-Arg-Glu-Asn*-Gly.

Thermolysin digestion of ChG-7 followed by fractionation of the digest by paper chromatography yielded a glycopeptide fraction and two peptides. Peptide Th-1, consisting of equimolar amounts of leucine and asparagine, proved to possess the sequence Leu-Asn and, therefore, represents the amino terminus of ChG-7. The other peptide (Th-2) was Sakaguchi-positive and consisted of equal amounts of serine, arginine, tyrosine, and valine plus isoleucine. Three direct Edman steps established the sequence to be

$$\underset{\text{Ile}}{\text{Val}}\text{-Ser-Arg-Tyr}$$

Hence, Th-2, which represents a mixture of *two variant peptides,* formed the carboxyl terminus of ChG-7. The glycopeptide fraction, because of the specificity of thermolysin, consisted of two components which differed in composition in that one possessed an additional amino-terminal asparagine residue (Th-G2). Direct Edman degradation afforded at each step a pair of "overlapping" amino acids. The sequences were found to be Asn-Val-Gln-X-Glu-Y-Gly-Thr and Val-Gln-X-Glu-Y-Gly-Thr. Based on these data the sequences of ChG-7 may be written as follows.



Peptide Ch-6 (Residues 92–98). This Pauly-positive peptide derived from fraction 7 (Figure 2) contained one glutaminyl residue. Direct Edman degradation yielded five steps: Val-Gly-Gly-Gln-Glu. Degradation with carboxypeptidase A for 4 hr afforded phenylalanine (1.00), histidine (0.20), and glutamic acid (0.10), and after 15 hr the corresponding values were 1.00, 0.80, and 0.30 respectively. With this evidence the sequence of Ch-6 was established as follows: Val-Gly-Gly-Gln-Glu-His-Phe. The sequence of this peptide was confirmed by Dr. R. A. Laursen utilizing the solid-phase peptide sequencer.

Peptide Ch-7 (Residues 99–101). Two steps of the subtractive Edman of this peptide which was derived from fraction 8 (Figure 2) established the sequence as Ala-His-Leu. Judging from the recovery of Ch-7, chymotrypsin cleaved the Leu-Leu bond [101–102) only to a small extent (see Ch-8).

Peptide Ch-8 (Residues 99–102). This peptide isolated from fraction 10 (Figure 2) afforded on digestion with carboxypeptidase A for 6 hr 2 mol of leucine and small amounts of histidine (0.2) and alanine (0.04). Three steps of the subtractive Edman technique established the sequence of this peptide as Ala-His-Leu-Leu. The recovery of Ch-8 confirms that chymotrypsin splits primarily the Leu-Ile bond.

Peptide Ch-9 (Residues 92–102). This peptide (fraction 9, Figure 2) represents a *variant peptide* to Ch-8 with Val in position 98. This slightly basic, Pauly- and Sakaguchi-positive peptide was digested with carboxypeptidase A for different periods of time, After 30 min the respective molar amounts of leucine, histidine, alanine, and valine were 2.0, 0.5, 0.1, and 0.1, after 2 hr 2.0, 1.0, 0.6, and 0.5, and after 20 hr 2.0, 1.5, 1.0, and 1.1, indicating the carboxyl-terminal sequence to be -His-(Val,Ala)-His-Leu-Leu. Analysis by paper chromatography in BAW of a total aminopeptidase M hydrolysate of Ch-9 revealed the absence of glutamine. Eight steps of the subtractive Edman established the following sequence: Glx-Gly-Gly-Arg-Glx-His-Val-Ala. To confirm this sequence, Ch-9 was digested with trypsin yielding two peptides whose composition and carboxyl-terminal amino acids were in agreement with the proposed sequence. Direct Edman degradation of the carboxyl-terminal tryptic peptide revealed the following amino-terminal sequence: Glu-His-Val-Ala-His. (The second and fifth steps yielded a strongly Pauly-positive, water-soluble phenylthiohydantoin compound indicating the presence of histidine in these two positions.) This sequence is again in agreement with the structure proposed for the original peptide: Glu-Gly-Gly-Arg-Glu-His-Val-Ala-His-Leu-Leu.

This peptide represents a *variant* of the two peptides Ch-6

plus Ch-8 and possesses *three amino acid substitutions*,[4] namely, a Glu for a Val in position 92, an Arg for a Gln in position 95, and a Val for a Phe in position 98. For this reason Ch-9 was not cleaved to a significant extent by chymotrypsin, but split by trypsin.

Peptide Ch-10 (Residues 103–110). Degradation by the direct and subtractive Edman procedures of this peptide isolated from fraction 13 (Figure 2) established the first five amino-terminal amino acids: Ile-Leu-Arg-Asp-Thr. Particular care was taken to ascertain that isoleucine was the amino terminus of this peptide. Analysis of a total enzymatic hydrolysate demonstrated the absence of asparagine. From a tryptic digest of Ch-10, a peptide composed of equimolar amounts of threonine and tyrosine, a tripeptide containing equimolar amounts of isoleucine, leucine, and arginine, and a second tripeptide consisting of aspartic acid, threonine, and lysine in equimolar amounts were isolated. One step of the subtractive Edman confirmed the sequence of the dipeptide and two steps that of the lysine-containing peptide. These data indicate that the sequence of Ch-10 is as follows: Ile-Leu-Arg-Asp-Thr-Lys-Thr-Tyr.

Peptide Ch-11 (Residues 104–110). This Sakaguchi-positive peptide isolated from fraction 11 (Figure 2) yielded on degradation by the subtractive Edman procedure the following partial sequence: Leu-Arg-Asp-Thr-Lys-Thr. Analysis of carboxypeptidase A digests of this peptide withdrawn at different times of incubation established the carboxyl-terminal sequence to be -Lys-Thr-Leu. Hence, the sequence of this peptide is Leu-Arg-Asp-Thr-Lys-Thr-Leu. This peptide represents a variant to Ch-10 and reveals *one amino acid substitution*, namely, a leucine for a tyrosine residue in position 110.

Peptide Ch-12 (Residues 105–110). This basic peptide isolated from fraction 4 (Figure 2) yielded on short digestion with carboxypeptidase A (20 min) tyrosine (0.90) and smaller amounts of threonine (0.22) and on a longer period of incubation (4 hr) tyrosine and threonine in molar quantities and traces of lysine. Five steps of the subtractive Edman established unambiguously the sequence of this peptide: Arg-Asp-Thr-Lys-Thr-Tyr.

Residue 111 (methionine) could not be determined from the study of the chymotryptic peptides, but was unambiguously established during the investigation of the corresponding pair of tryptic peptides (T-10 and T-10, variant).

Amino Acid Sequences of Certain Peptic Peptides and Glycopeptides

Peptide P-1 (residues 19–25) proved to be identical with Ch-2. A *variant peptide* (P-1, V; residues 19–25) was also elucidated and confirmed the amino acid replacement of an arginine for a glutamine residue in position 20. Peptide P-2 (residues 27–32), which was shown to possess an amino-terminal tyrosine (residue 27) and had a carboxyl-terminal alanine as a *replacement* for phenylalanine in position 32 of Ch-3, was otherwise identical with the latter peptide. Additional peptides had sequences of residues 44–47 (P-3), 44–48 (P-4), and 48–46 (P-5) and corresponded to sections of ChG-3. Peptide P-3 had an *amino acid substitution* in position 47 (alanine for threonine). Two peptides overlapping with certain chymotryptic peptides possessed the sequences of residues 91–99 (P-7) and 102–104 (P-9), respectively. Two further peptides were characterized by their sequences from residues 62 to 64 (P-6) and from 99 to 101 (P-8).

The peptic glycopeptides, because of the particular amino acid sequences of α_1-acid glycoprotein and because of the specificity of pepsin, had somewhat shorter polypeptide chains than the chymotryptic glycopeptides. The glycopeptide with the heteropolysaccharide unit 1 reached from residues 9 through 18 (PG-1), that with the unit 2 from residues 32 through 43 (PG-2), representing an important overlapping peptide, that with unit 3 from residues 50 through 58 (PG-3), and that with unit 5 from residues 80 through 90 (PG-5), agreeing with the sequences of the chymotryptic glycopeptides.

TABLE V: Amino Acid Composition of CNBr-I of α_1-Acid Glycoprotein.

Amino Acid	Compiled from Established Sequence: No. of Full Residues without Substn	No. of Residues with Substn[a]	Total No. of Residues	Av of 3 Amino Acid Analyses
CM-Cys	2	*b*	2	1.62[c]
Asx	12	*b*	12	12.00
Thr	10	3	11.5	11.66[d]
Ser	3	2	4	3.63[d]
Glx	14	3	15.5	15.15
Pro	4	*b*	4	3.60
Gly	4	*b*	4	4.25
Ala	6	2	7	6.10
Val	4	3	5.5	5.62
Ile	7	2	8	7.63[e]
Leu	8	1	8.5	8.91[e]
Tyr	7	1	7.5	7.59[f]
Phe	5	3	6.5	7.17
Lys	4	*b*	4	5.33
His	1	1	1.5	2.08
Arg	6	3	7.5	6.25
Trp	1	*b*	1	(1)
Hser	1	*b*	1	(1)
Total	99	24	111	110.59

[a] Each substitution was counted as half a residue. [b] No substitutions were observed. [c] Extrapolated. [d] Corrected for destruction during hydrolysis. [e] The sum of Ile and Leu equals the expected value. [f] Phenol was present during hydrolysis (Howard and Pierce, 1969).

Discussion

The amino acid composition of the amino-terminal CNBr fragment of α_1-acid glycoprotein proved to be very unusual as no integers were obtained (Table V). It was not until the large number of hitherto unknown amino acid substitutions of this protein discussed below were elucidated that these data could be interpreted. After these replacements were taken into consideration by calculating the amino acid composition from the elucidated sequence, a surprisingly good agreement with the original analysis was obtained (Table V), even though each replacement was arbitrarily counted as half a residue. However, the comparison between the calculated and the

[4] These amino acid replacements like most others were confirmed by the structures of the corresponding tryptic peptides (T-5, T-6, T-7, and T-8) (see also Table VI).

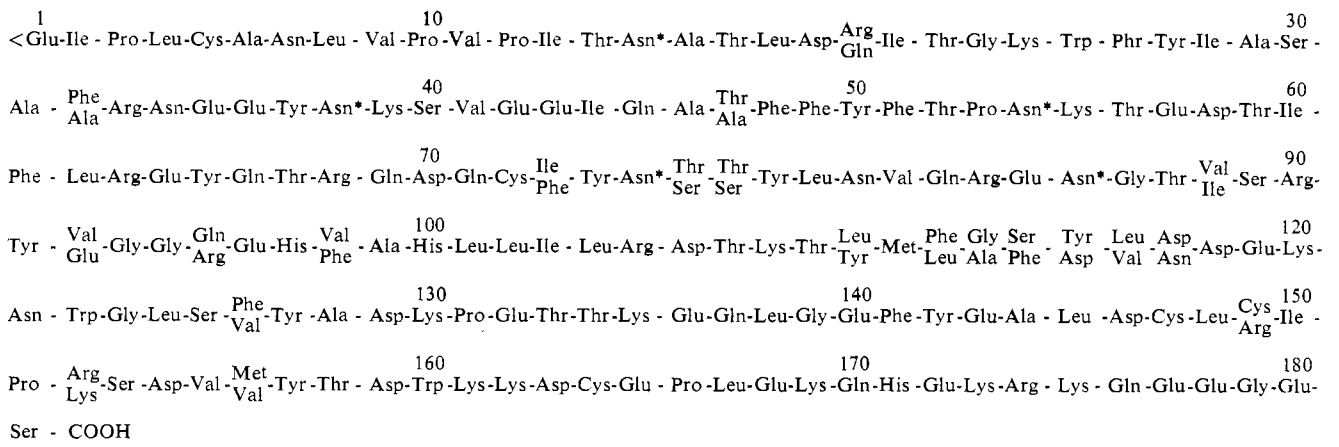


FIGURE 6: The complete linear amino acid sequence of α_1-acid glycoprotein. The five heteropolysaccharide units are linked N-glycosidically to the asparagine residues that are marked with asterisks.

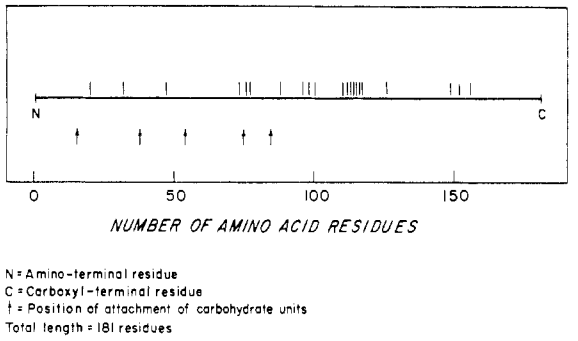


FIGURE 7: Schematic presentation of the amino acid substitutions of α_1-acid glycoprotein. The short vertical lines above the horizontal line which represents the polypeptide chain of the protein point to the positions of these replacements. The vertical arrows below the horizontal line point to the positions of attachment of the five heteropolysaccharide units of this α_1-globulin.

experimentally obtained composition seems to indicate that the degree of substitution of certain amino acids is considerably different from 50%.[5]

The amino acid sequence of this CNBr fragment was unambiguously established on the basis of the sequences of the chymotryptic, tryptic, and peptic peptides and glycopeptides.[6] It should be noted that the amino acid sequence of CNBr-I together with that of the carboxyl-terminal CNBr fragment (Ikenaka *et al.*, 1972) establish the complete linear amino acid sequence of α_1-acid glycoprotein (Figure 6) and describe for the first time the sequence of a protein with five heteropolysaccharide units.

In view of the many essential biological functions exhibited by glycoproteins, especially those of the cell membranes in the normal and pathological states (Schmid, 1972; Jamieson and Greenwalt, 1971), three points of the present study that appear pertinent for future studies of complex conjugated proteins using an automated sequence apparatus are mentioned in the following. Proteins with pyrrolidonecarboxylic acid can only be sequenced after this residue is cleaved enzymatically. L-(1-Tosylamido-2-phenyl)ethyl chloromethyl ketone trypsin hydrolyzes extremely rapidly certain aromatic (-Phe-Phe-Tyr-Phe-) and very rapidly certain highly aliphatic (-Leu-Leu-Ile-Leu-) regions. This enzyme will effect certain unexpected cleavages such as the Leu-CM-Cys bond leading to the consistent formation of the amino-terminal tetrapeptide.

The amino acid sequence of CNBr-I proved very hydrophobic in nature as it consists of a surprisingly large number of aromatic and highly aliphatic residues, in contrast to the hydrophilic properties of CNBr-II [Ikenaka *et al.*, 1972). However, CNBr-I with the five carbohydrate units of the protein possesses a large number of aliphatic hydroxyl and carboxylic groups (ten sialyl residues) which are located on the surface of the molecule and thereby render this portion of the protein also hydrophilic. Thus, this asymmetric distribution of the sialyl residues results in a relatively even distribution of the electrostatic charges along the whole polypeptide chain of the molecule and probably explains many of the physicochemical properties of α_1-acid glycoprotein (Jeanloz, 1972).

The distances between the five heteropolysaccharide units of this protein, whose positions of attachment to the polypeptide chain were determined in the present study, can now be evaluated and, in terms of the number of amino acid residues between two consecutive carbohydrate–polypeptide bonds, were found to vary considerably being 22, 15, 20, and 9, respectively. The actual distances between these carbohydrate units within the three-dimensional structure of the protein and the relative directions of these prosthetic groups on the surface of the protein molecule will be revealed by X-ray diffraction pattern analysis.

The number of amino acid substitutions found in the amino- and carboxyl-terminal (Emura *et al.*, 1971) CNBr fragments of α_1-acid glycoprotein (summarized in Table VI) is unusually high (11 and 10, respectively), *i.e.*, 21 out of 181 residues or 12%. Some of these substitutions may represent genetic markers which will be evaluated in a future immunochemical study. No other protein, excepting the immunoglobulins, is known to possess such a high number of amino acid replacements. This high frequency of amino acid substitutions suggested a relationship between α_1-acid glycoprotein and the immunoglobulins as discussed below. Of further interest is the

[5] It should be realized that amino acid substitutions present in very small percentages escaped detection, while for amino acid replacements present in somewhat higher percentages only indications were obtained. (*E.g.*, at position 41, a small amount of isoleucine was obtained in addition to valine. At position 101 a small amount of threonine was found in addition to leucine, the former being present as the variant peptide: Ala-His-Thr.)

[6] In their preliminary study on α_1-acid glycoprotein, Kitamura and Yamashina (1972) described the limited amino acid sequences of four tryptic glycopeptides but did not report on the peptides that connect these glycopeptides with each other. These sequences are in essential agreement with those reported in the present paper.

TABLE VI: Amino Acid Substitutions of Pooled α_1-Acid Glycoprotein.

Position in Sequence	Amino Acid Substn[a]	No. of Base Changes in Codon	Symbol of the Peptides and Variant Peptides Isolated from a Digest with[b]		
			Chymotrypsin	Pepsin	Trypsin
20	Arg-Gln	1	ChG-2 (ChG-2, V)	P-1 (P-1, V)	TG-1,[c] TG-2[c]
32	Phe-Ala	2	Ch-3	(P-2)	T-2, T-3
47	Thr-Ala	1	ChG-3	(P-3)	TG-4
73	Ile-Phe	1	Ch-5		
76	Thr-Ser	1	ChG-6[c]		TG-9 (TG-9, V)
77	Thr-Ser	1	ChG-6		
88	Ile-Val	1	ChG-1[c]		TG-10[c]
92	Val-Glu	1	Ch-6 (Ch-9)		T-5 (T-6), (T-7)
95	Gln-Arg	1	Ch-6 (Ch-9)		
98	Val-Phe	1	Ch-6 (Ch-9)		T-6 (T-8)[c]
110	Leu-Tyr	2	Ch-10, Ch-12 (Ch-11)		T-10 (T-10, V)
112	Phe-Leu	1			
113	Gly-Ala	1			
114	Ser-Phe	1			
115	Tyr-Asp				
116	Leu-Val	1	amino acid substitutions in positions 110–156 are present in CNBr-III[d]		
117	Asp-Asn	1			
126	Phe-Val	1			
149	Cys-Arg	1			
152	Arg-Lys	1			
156	Met-Val	1			

[a] The amino acids which were found in the variant peptides are the second of each set of amino acids. [b] The variant peptides are in parentheses. [c] This peptide represents a mixture of the normal and variant peptides. [d] Emura *et al.*, 1971.

distribution of the amino acid substitutions on the polypeptide chain on this glycoprotein appearing somewhat symmetrical with a large number in the middle of the polypeptide chain (Figure 7). Seven of these substitutions were found in the segment that is homologous with the α chain of haptoglobin (Ikenaka *et al.*, 1972). Regarding the minimum number of mutations required to explain these substitutions, it should be noted that 19 of these amino acid replacements may each have involved a single base change in their codons, whereas two required two base changes each in their codons (Table VI). In these replacements, Arg is involved four times and Thr, Ala, Leu, Ile, Ser, Gln, and Asp each twice. The following pairs of amino acid substitutions occurred each twice: Phe-Val, Ser-Thr, and Arg-Gln. Excepting tryptophan and proline, all amino acids are involved in these changes. Further studies are required to elucidate which of these amino acid substitutions are located on the same variant molecule.

The presence of these amino acid replacements in α_1-acid glycoprotein may explain our earlier finding (Schmid *et al.*, 1968) that desialyzed, pooled α_1-acid glycoprotein could be resolved by chromatography on DEAE-cellulose into seven variant protein preparations. These preparations revealed varying differences in their electrophoretic mobilities (Schmid *et al.*, 1968) indicating different electrostatic net charges of the variant protein molecules and, thus, different numbers of amino acid replacements. In support of this resolution of α_1-acid glycoprotein variant molecules is the finding that, using ion exchange chromatography on DEAE-cellulose (Figure 4F), large glycopeptides that differ only in the content of one phenylalanine or tyrosine residue could be separated from each other.

In view of the earlier reported homology between the carboxyl-terminal CNBr fragment of this plasma glycoprotein and a constant region of the H chain of IgG (Ikenaka *et al.*, 1972), and in view of the multiple amino acid substitutions of the amino-terminal CNBr fragment of this plasma globulin, CNBr-I was also analyzed for homology with the immunoglobulins. It was found that the direct homology (percentage of identical residues) of the 43 amino-terminal residues of CNBr-I with the corresponding amino-terminal region of the κ-type L chain of human IgG was 27% (Figure 8). According to Dayhoff (1972) homology values above 23% are significant. This homology of CNBr-I appeared even more significant when one considers that the mentioned homologous section of the L chain represents part of the variable segment and, furthermore, includes one of the hypervariable areas. The extent of homology, when single point mutations were allowed, increased to 80%. Evaluation by the procedure of Fitch

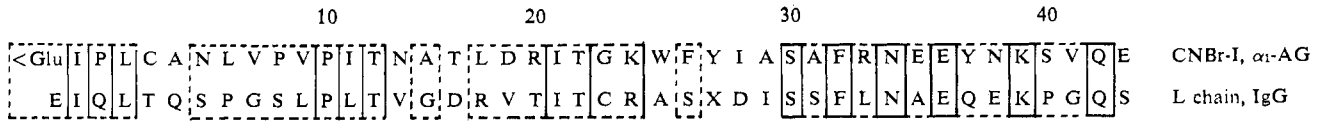


FIGURE 8: The homology between the amino-terminal 43-residue segment of α_1-acid glycoprotein and the amino-terminal segment of a κ-type L chain. For this presentation the single-letter code for amino acids is used (Dayhoff, 1972). Identical residues are boxed in and the residues which require a single point mutation for identity are in boxes with broken frames.

(1966a,b, 1969) also revealed this homology to be significant (Schmid *et al.*, unpublished results). From these observations it is hypothesized that a possible relationship exists between α_1-acid glycoprotein and the ancestral immunoglobulin (Hill *et al.*, 1966).

Acknowledgments

The authors are very grateful to Dr. R. F. Doolittle, University of San Diego, for the gift of lyophilized *Pseudomonas fluorescens* cells from which pyrrolidonecarboxylylpeptidase was isolated, to Professor M. Funatsu and Dr. M. Ishiguro, Kyushu University, Fukuoka, Japan, and Mr. G. Crombie and Dr. C. Franzblau for their kindness in carrying out innumerable amino acid analyses, to Mr. Marcus E. Schmid, The Johns Hopkins University, for evaluating the homology between α_1-acid glycoprotein and the immunoglobulin by the technique of Fitch, to Dr. R. A. Laursen for elucidating the amino acid sequence of certain peptides with his solid phase peptide sequencer, and to Mr. L. Larsen, Institute of Laboratories, Massachusetts Department of Health, Jamaica Plain, Mass., for his generous supply of Cohn fraction VI from which α_1-acid glycoprotein was isolated. We also wish to acknowledge the technical assistance of Miss Hedy Hediger and Mrs. K. Hunziker.

References

Anderson, B., Seno, N., Sampson, P., Riley, J. G., Hoffman, P., and Meyer, K. (1964), *J. Biol. Chem. 239*, 2716.

Bencze, W. L., and Schmid, K. (1957), *Anal. Chem. 29*, 1193.

Blackburn, S. (1968), Amino Acid Determination—Methods and Techniques, New York, N. Y., Marcel Dekker.

Bürgi, W., and Schmid, K. (1961), *J. Biol. Chem. 236*, 1066.

Crestfield, A. M., Moore, S., and Stein, W. H. (1963), *J. Biol. Chem. 238*, 622.

Dayhoff, O. (1972), Atlas of Protein Sequence and Structure, Silver Spring, Md., National Biomedical Research Foundation.

Doolittle, R. F. (1972), *Methods Enzymol. 25B*, 231.

Elzinga, M., Lai, C. Y., and Hirs, C. H. W. (1965), *Arch. Biochem. Biophys. 123*, 353.

Emura, J., Ikenaka, T., Collins, J. H., and Schmid, K. (1971), *J. Biol. Chem. 246*, 7821.

Fitch, W. M. (1966a), *J. Mol. Biol. 16*, 9.

Fitch, W. M. (1966b), *J. Mol. Biol. 16*, 17.

Fitch, W. M. (1969), *J. Mol. Biol. 49*, 1.

Goldberger, R. F., and Anfinsen, C. B. (1962), *Biochemistry 1*, 401.

Gray, W. R. (1967), *Methods Enzymol. 11*, 139.

Hill, R. L., Delaney, R., Fellows, R. E., and Lebowitz, H. E. (1966), *Proc. Nat. Acad. Sci. U. S. 56*, 1762.

Howard, S. M., and Piece, J. G. (1969), *J. Biol. Chem. 244*, 6468.

Ikenaka, T., Bammerlin, H., Kaufmann, H., and Schmid, K. (1966), *J. Biol. Chem. 241*, 5560.

Ikenaka, T., Ishiguro, M., Emura, J., Kaufmann, H., Isemura, S., Bauer, W., and Schmid, K. (1972), *Biochemistry 11*, 3817.

Ishiguro, M., Yamamoto, A., and Schmid, K. (1969), *Biochim. Biophys. Acta 194*, 91.

Jackson, R. L., and Hirs, C. H. W. (1970), *J. Biol. Chem. 245*, 624.

Jamieson, G. A., and Greenwalt, T. J. (1971), Glycoproteins of Blood Cells and Plasma, Philadelphia, Pa., J. B. Lippincott Co.

Jeanloz, R. W. (1972), *in* Glycoproteins, Vol. 5, Gottschalk, A., Ed., Amsterdam, Elsevier, pp 564–611.

Kitamura, K., and Yamashina, I. (1972), *J. Biochem. (Tokyo) 71*, 65.

Laursen, R. A. (1971), *Eur. J. Biochem. 20*, 89.

Nimberg, R. B., Motoyama, T., and Schmid, K. (1971), *J. Biol. Chem. 246*, 5817.

Schmid, K. (1972), *Chimia 26*, 405.

Schmid, K., Kaufmann, H., Isemura, S., Bauer, F., Emura, J., Motoyama, I., Ishiguro, M., and Nanno, S. (1971a), *Fed. Proc., Fed. Amer. Soc. Exp. Biol. 30*, 1413 Abstr.

Schmid, K., Kaufmann, H., Isemura, S., Bauer, F., Emura, J., Motoyama, T., Ishiguro, M., and Nanno, S. (1971b), *Biochem. Biophys. Res. Commun. 42*, 280.

Schmid, K., Kaufmann, H., Isemura, S., Bauer, F., Emura, J., Motoyama, T., Ishiguro, M., and Nanno, S. (1972), *Experientia 28*, 24.

Schmid, K., Okuyama, T., and Kaufmann, H. (1968), *Biochim. Biophys. Acta 154*, 562.

Schmid, K., Polis, A., Hunziker, K., Fricke, R., and Yayoshi, M. (1967), *Biochem. J. 104*, 361.

Schroeder, W. A. (1972), *Methods Enzymol. 25B*, 203.

Spiro, R. G. (1966), *Methods Enzymol. 8*, 3.

Svennerholm, L. (1963), *Methods Enzymol. 6*, 459.

Szewczuk, A., and Kwiatkowska, J. (1970), *Eur. J. Biochem. 15*, 92.

Yemm, E. W., and Cocking, E. C. (1955), *Analyst (London) 80*, 209.