Isolation and Partial Characterization of the Cyanogen Bromide Fragments of α_1-Acid Glycoprotein and the Elucidation of the Amino Acid Sequence of the Carboxyl-Terminal Cyanogen Bromide Fragment†

Tokuji Ikenaka, Masatsune Ishiguro, Junji Emura, Hugo Kaufmann, Satoko Isemura, Wilfred Bauer, and Karl Schmid*

ABSTRACT: For the elucidation of the amino acid sequence of α_1-acid glycoprotein (orosomucoid), a globulin of normal human plasma, this protein was first subjected to CNBr cleavage. After subsequent reduction and carboxymethylation four fractions were obtained by gel filtration through a long Sephadex G-100 column. Fraction I contained the total carbohydrate moiety of the protein and represents the amino-terminal fragment. Fraction II which consisted of 70 amino acid residues was shown to represent the carboxyl terminus. The amino acid sequence of fragment II was established. Because of its large number of lysine residues this fragment was digested with chymotrypsin on the one hand and, on the other hand, was citraconylated and then digested with trypsin in order to obtain overlapping peptides. The sequence of this fragment proved to be as follows: Leu-Ala-Phe-Asp-Val-Asn-Asp-Glu-Lys-Asn-Trp-Gly-Leu-Ser-Val-Tyr-Ala-Asp-Lys-Pro-Glu-Thr-Thr-Lys-Glu-Gln-Leu-Gly-Gln-Phe-Tyr-Glu-Ala-Leu-Asp-Cys-Leu-Arg-Ile-Pro-Lys-Ser-Asp-Val-Val-Tyr-Thr-Asp-Trp-Lys-Lys-Asp-Cys-Glu-Pro-Leu-Glu-Lys-Gln-His-Glu-Lys-Arg-Lys-Gln-Glu-Glu-Gly-Glu-Ser. It should be noted that the amino-terminal 22-residue segment of fragment II possessed a high degree of homology with the α chain of haptoglobin. Even more striking was the observation that the remainder of this fragment was highly homologous with the constant region of the H chain of the immunoglobulins. Thus, α_1-acid glycoprotein is the first single-chain protein that possesses sequence similarities with two other plasma proteins.

α_1-Acid glycoprotein (orosomucoid) is probably one of the most extensively characterized glycoproteins of human plasma (Jeanloz, 1966), especially with regard to its carbohydrate moiety (Wagh *et al.*, 1969) which accounts for 40% of the protein. While certain aspects of the amino acid and monosaccharide sequences of this protein have been reported (Jeanloz, 1966), the progress in elucidating its complete primary structure has been hampered by (1) the pronounced resistance[1] of the native protein toward digestion by specific proteases (Bourrillon and Meyer, 1967; Yamashina, 1965) and (2) the high tendency of the protein to form, upon certain chemical modifications, aggregates that are again resistant to enzymatic digestion.

In the present study these problems were circumvented by the highly specific fragmentation of α_1-acid glycoprotein with cyanogen bromide (CNBr).[2] CNBr degradation (Spande *et al.*, 1970), as a first step for the elucidation of the amino acid sequence, appeared particularly suited as this protein contains a very small number of methionine residues. In this paper the amino acid sequence of the carboxyl-terminal CNBr fragment of α_1-acid glycoprotein which is devoid of sugar and consists of 70 amino acid residues is presented.

Materials and Methods

α_1-Acid glycoprotein was isolated from the supernatant solution of Cohn fraction V of pooled normal human plasma (Bürgi and Schmid, 1961). The homogeneity of this globulin was established by several criteria of purity including amino- and carboxyl-terminal amino acid analyses (Jeanloz, 1966; Ikenaka *et al.*, 1966). The molecular weight of this protein was assumed to be 40,000.[3] Amino acid analysis indicated that this globulin contains approximately 1.3 methionyl residues (Table II, see below).

Trypsin (TPCK-treated, three-times crystallized), *α-chymotrypsin* (three-times crystallized), *DFP-CPA*, and *DFP-CPB* were purchased from Worthington Biochemical Corp., Freehold, N. J., while *CPC* was a gift from Dr. H. Zuber, Zürich, Switzerland (Zuber, 1968). *Thermolysin* was obtained from Calbiochem, Los Angeles, Calif., and *AP-M* from Rohm and Haas through Henley & Co., N. Y. All other reagents were of analytical grade and utilized without further purification.

Preparation of the CNBr Fragments. α_1-Acid glycoprotein (2.2 g) was dissolved in a stoppered 1-l. round-bottom flask by the addition of 20 ml of 70% formic acid. CNBr (Eastman, Rochester, N. Y.) (2 g) was added and the resulting solution

† From the Department of Biochemistry, Boston University School of Medicine, Boston, Massachusetts 02118. *Received April 28, 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. A preliminary account of this work was reported earlier (Ikenaka *et al.*, 1971).

[1] R. B. Nimberg and K. Schmid, unpublished data.

[2] Abbreviations used are: CNBr, cyanogen bromide; DFP, diisopropyl phosphorofluoridate; Dnp, dinitrophenyl; Gdn·HCl guanidine hydrochloride; BAW, 1-butanol–acetic acid–water (200:30:75 or 3:1:1); BAWP, 1-butanol–acetic acid–water–pyridine (60:12:48:40); glpc, gas–liquid partition chromatography; pc, paper chromatography; hve, high-voltage electrophoresis; tlc, thin-layer chromatography; PTH phenylthiohydantoin; CPA, carboxypeptidase A; CPB, carboxypeptidase B; CPC carboxypeptidase C; AP-M, aminopeptidase M; NANA, *N*-acetylneuraminic acid.

[3] K. Kawahara, R. B. Nimberg, and K. Schmid, unpublished data.

was stirred under nitrogen at room temperature for 20 hr (Steers *et al.*, 1965), diluted with 200 ml of water, and lyophilized. The extent of conversion of the methionine to homoserine residues was determined by an automatic amino acid analyzer, after total acid hydrolysis of the modified glycoprotein. For the conversion of homoserine lactone to homoserine the hydrolysates were treated with 1 *M* piperidine at 40° for 1 hr and applied immediately to the mentioned analyzer. The newly formed amino-terminal amino acids were identified and quantitatively determined as Dnp derivatives using two-dimensional paper chromatography (Fraenkel-Conrat *et al.*, 1955) and as PTH-amino acids (Iwanaga *et al.*, 1969).

For its reduction the CNBr-treated α_1-acid glycoprotein which was insoluble in water, in contrast to the native protein, was dissolved in 20 ml of 0.1 M Tris-HCl buffer (pH 8.6), containing 11.5 g of Gdn·HCl (6 M) and 40 mg of Na_2-EDTA. The pH of the solution was adjusted to 8.6 and 1.0 ml of β-mercaptoethanol was added. The resulting solution was stirred at room temperature for 7 hr (Ishiguro *et al.*, 1969). For the subsequent alkylation of the cleaved CNBr fragments, 2.7 g of iodoacetate dissolved in 2.0 ml of 2 N NaOH was added in small aliquots, and the pH of the reaction mixture was kept at 8 by addition of 1 N NaOH. The separation of the alkylated CNBr fragments was achieved by gel filtration.

Procedures Used for the Elucidation of the Amino Acid Sequence

Chymotryptic, Tryptic, and Peptic Hydrolysis. To a solution of 100 mg of fragment II in 10 ml of 0.1 M pH 8.0 $(NH_4)_2$-CO_3–$NH_4(HCO_3)$ buffer, 0.5 mg of trypsin or chymotrypsin was added and the digestion was carried out at 38° for 3 hr. The pH of the digest was kept constant by addition of 0.1 N NaOH with the aid of a Radiometer pH-Stat (Model No. TTTlc). This reaction was subsequently terminated by lowering the pH to 4.0 by addition of acetic acid, and the digest subsequently lyophilized. The peptic digestion (3% enzyme) was performed at pH 2.0.

Citraconylation and Subsequent Tryptic Digestion. Fragment II (100 mg) was dissolved in water (7 ml) and the pH was adjusted to 8.0. Over a period of 30 min seven aliquots (100 mg) of citraconic acid anhydride (Eastman) were added (Dixon and Perham, 1968). The pH was maintained at 8.0 by addition of 5 N NaOH. Subsequently, the modified fragment was desalted by gel filtration through a Sephadex G-25 column (2.4 × 28 cm) equilibrated with 0.5% ammonium carbonate buffer (pH 8.0), followed by tryptic digestion and lyophilization. The resulting tryptic peptide mixture was dissolved in 60% acetic acid and allowed to stand at 24° overnight to cleave the blocking agent.

Separation of the Chymotryptic, Tryptic, and Peptic Peptides. A lyophilized enzymatic digest dissolved in 2.0 ml of 0.2 M acetic acid was fractionated by gel filtration (see Figure 5). Appropriate fractions as indicated in this figure were subsequently pooled, dried, and redissolved in 0.5 ml of water, 5% acetic acid, or 5% pyridine. All peptide fractions to be studied, except those which contained basic peptides, were purified by ion-exchange chromatography on Dowex 50-X2 and Dowex 1W-X2 (Schroeder, 1967). Purification of the basic peptides was achieved by preparative high-voltage electrophoresis at pH 1.9 in 5% formic acid or at pH 3.5 or 6.4 in pyridine–acetate buffer (Ryle *et al.*, 1955) utilizing Whatman No. 3MM. Moreover, gel filtration through Sephadexes, preparative chromatography (PC) in BAW and BAWP, and chromatography on DEAE-cellulose were also used, depending on the results of preliminary experiments.

Purity of Peptides. The homogeneity of the peptides was established by descending chromatography on Whatman No. 1MM paper in BAW and/or BAWP and by hve on Whatman No. 3MM paper utilizing a water-cooled flat-plate system (Model 1 FP Savant Instruments, Inc., Hicksville, N. Y.). A potential of 50–55 V/cm was applied. Two of the following buffers were employed: (a) pH 1.9, (b) pH 3.4, and (c) pH 6.4 (Ryle *et al.*, 1955). Sodium borate buffer (pH 9.2, 0.05 M) was also utilized. After chromatography or hve, the paper was stained with ninhydrin, followed by the chlorine–starch reagent using *tert*-butyl hypochlorite. In order to detect the tryptophan- and arginine-containing peptides, the papers were stained with the Ehrlich and Sakaguchi reagents (Canfield and Anfinsen, 1963), respectively. As judged by these criteria the peptides were at least 95% pure. Determination of the amino-terminal amino acid (see below) fulfilled an additional criteria of purity. Appropriate aliquots of these peptide solutions were used for determining the amino acid composition and for calculating the content of the peptides.

Amino Acid Composition. Prior to amino acid analysis all peptides were hydrolyzed with redistilled HCl in evacuated and sealed tubings for 24 hr and after removal of the HCl, applied to a Technicon, Hitachi Model KLA-3 or Beckman Model C amino acid analyzer. The composition of small peptides was also determined by glpc (Gehrke and Stalling, 1970) which gave results[4] identical with those obtained by the classical ion-exchange AutoAnalyzer. For this glpc procedure Tabsorb purchased from Regis Co., Chicago, Ill., was utilized as stationary phase. For the determination of asparagine and glutamine as well as tryptophan the peptides were hydrolyzed with AP-M. Depending on the amino acid composition of the peptides, asparagine and glutamine could be determined directly by hve at pH 6.4 or pc in BAW followed by staining with the Cd–ninhydrin reagent (Canfield and Anfinsen, 1963) and quantitative determination of each amino acid. In general, however, the contents of these amides were calculated from the total content of aspartic and glutamic acids as determined from acid hydrolysates minus the content of aspartic and glutamic acid established from the analysis of the enzymatic hydrolysates. Methionine was determined as methionine sulfone after oxidation with performic acid of the glycoprotein and subsequent hydrolysis (Moore, 1963). This technique afforded simultaneous determination of cysteine as cysteic acid.

Terminal Amino Acid Analysis. The direct Edman procedure (0.5 μM of a peptide) (Iwanaga *et al.*, 1969) was employed for the determination of the amino-terminal amino acids of fragment II and the peptides derived thereof. For the identification of amino acids cleaved by CPA, CPB (Hirs, 1967), CPC or AP-M, pc in BAW or BAWP, hve at 1.9 or 6.4, and/or glpc (Gehrke and Stalling, 1970) were used depending on the released amino acids. The amino acids released by CPA, CPB, and CPC are indicated in Results by arrows (←) under which their molar recoveries are given. The degradation technique of Cromwell and Stark (1969) was utilized for the determination of the carboxyl-terminal amino acid of peptides which were not degraded by CPA or CPC.

During the earlier part of this study the amino-terminal amino acids were determined by the Dnp procedure (Frankel-Conrat *et al.*, 1955). When the native or CNBr-treated glyco-

[4] H. Hediger and K. Schmid, unpublished data.

TABLE I: Properties of CNBr-Treated α_1-Acid Glycoprotein and of the Four Fragments (I, II, III, and IV) Derived Therefrom.

Property	AG[a]	CNBr AG	I	II	III	IV
Yield, weight, %		97	68	14	3	3
Yield, molar, %		97	86	70	25	25
Estimated molecular weight	40,000	38,000	29,000	8000	5000	3000
Amino-terminal amino acids (mole/mole)	PCA (1.0)[b]	PCA (1.0)	PCA (1.0)[c]			
		Leu (0.5)[d]		Leu (1.0)		
		Phe (0.2)			Phe (1.0)	
		Tyr[e]				Tyr (0.9)[f]
Carboxyl-terminal amino acids (mole/mole)	Ser (1.0)[g]	Ser[h]		Ser (1.0)[i]		Ser (1.0)[i]
		Hse (1.3)	Hse (1.0)		Hse (1.0)	
		Carbohydrate Composition				
Neutral hexoses, %	16	16	20			
Hexosamine, %	15	15	19			
Sialic acid, %	11	8	10			
Fucose, %	1	ND	ND			

[a] Abbreviations used are: AG, α_1-acid glycoprotein; CNBr-α_1-AG, cyanogen bromide treated α_1-acid glycoprotein. [b] See Schmid (1954). The number in parentheses indicates the recovery of the terminal amino acid expressed in moles per mole of compound. [c] See Kaufmann and Schmid (1972). [d] Not corrected for losses. Done by both the dinitrophenylation and the direct Edman techniques. [e] Tyrosine was formed in minute quantities only. [f] Determined by the direct Edman procedure (Iwanaga *et al.*, 1969). [g] See Schmid *et al.* (1959). [h] This carboxyl-terminal serine was not determined. See also Schmid *et al.* (1959). [i] Determined by the procedure of Cromwell and Stark (1969).

protein or fragment I was analyzed a yellow artifact was always present in the final water phase containing the water-soluble Dnp compounds. This artifact, which had an R_F value in the *tert*-amyl alcohol–potassium phthalate buffer similar to that of Dnp-Arg and the same mobility on hve, was Sakaguchi negative.

Determination of the Amino Acid Sequences. The *subtractive Edman* degradation technique (Elzinga *et al.*, 1968) in conjuction with amino acid analyses and primarily the quantitative *direct Edman* (Iwanaga *et al.*, 1969) procedure were employed. For the latter technique, Beckman's sequential reagents and Analteck's tlc plates (silica gel G F 250 from Analteck, Newark, Del.) were utilized. The amino acids cleaved by this procedure are indicated in Results by arrows (→) under which their molar recovery is given. Threonine and serine indicated by (*) cannot be determined quantitatively by this procedure. These two PTH-amino acids exhibit a characteristic spectrum in the range from 260 to 300 nm. The aqueous layers of the Edman steps, when containing arginine or histidine, were concentrated and subjected to hve at pH 6.4. The migration rates of these two PTH-amino acids differ characteristically from each other. The presence of arginine was further demonstrated by the use of the Sakaguchi stain indicated in Results by (+). The partial reaction of PhSCN with the ϵ-amino groups and the subsequent incomplete hydrolysis led to decreasing recoveries of lysine observed by the subtractive Edman procedure. The technical details of the many additional procedures necessary for the determination of the amino acid sequence are described elsewhere (Hirs, 1967) and in the current literature (Landon *et al.*, 1971; Shinoda *et al.*, 1970; Huang and DeLange, 1971).

Nomenclature of Peptides. Chymotryptic peptides are designated with the letter C, tryptic peptides by T, peptic peptides by P, the citraconylated peptides by LC,[5] and the thermolytic peptides with Th. The chymotryptic peptides are numbered according to the sequence of their elution from the first Sephadex column and the tryptic peptides according to the sequence of their isolation.

Additional Analytical Procedures. Hexose was measured by the orcinol technique, hexosamine by the Morgan–Elson procedure, and sialic acid by the thiobarbituric acid method (Spiro, 1966). The content of tryptophan of CNBr fragments was determined spectrophotometrically according to the procedure of Bencze and Schmid (1957). For ultracentrifugal analyses a Spinco Model E ultracentrifuge equipped with an automatic temperature controller and schlieren and interference optics were used. The molecular weights of the larger CNBr fragments were estimated by the sedimentation equilibrium technique of Yphantis (1960). The partial specific volume was assumed to be 0.670 (ml/g) for fragment I (Popenoe and Drew, 1957) and 0.73 (ml/g) for fragment II.

Results

Isolation and Partial Characterization of the CNBr Fragments

Cleavage of α_1-Acid Glycoprotein with CNBr. The amino- and carboxyl-terminal amino acids of the CNBr-treated glycoprotein are listed in Table I. The carbohydrate composition of this modified glycoprotein (Table I) indicated that the content of the neutral and basic hexoses was not changed significantly. However, the content of NANA decreased from 11 to 8%, and the formed free NANA accounted indeed for the difference. On paper electrophoresis at pH 8.6, the modified glycoprotein migrated as a homogeneous component with a mobility corresponding to that of the α_1-globulins. On starch gel electrophoresis at the same pH in presence of 8 M urea a single band was also noted. Ultracentrifugal analysis revealed a single refractive index gradient which sedimented with a coefficient of 2.1 S at a concentration of 0.4%

[5] The abbreviation LC should indicate that the ϵ-amino group of the lysine residues is substituted by a citraconyl residue.

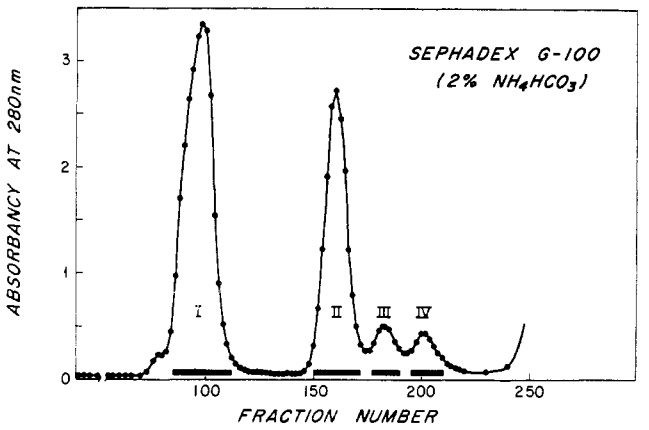

FIGURE 1: Fractionation by gel filtration of the CNBr-treated α_1-acid glycoprotein after reduction and carboxymethylation. The column size was 3.5 × 225 cm and the flow rate 24 ml/hr.

(the native protein at a concentration of 1% sedimented with a coefficient of 3.0 S). This observation suggested that all CNBr fragments were linked together by disulfide bonds as confirmed below.

The amino acid composition of the CNBr-treated glycoprotein revealed the lack of methionine and the presence of homoserine (Tables I and II). The recovery of homoserine of 1.4 moles/mole of modified glycoprotein appeared to be in agreement with the molar recovery of fragments II, III, and IV and, in the light of the results described below, can probably be explained as follows. The cleavage of one methionine residue led to the formation of approximately 1 mole of the amino- (I) and 0.70 mole of the carboxyl-terminal (II) fragment of the protein. The two other fragments (III and IV) were obtained in molar yields of 0.25 and, as will be shown later, were derived from a methionine variant of fragment II present in the pooled glycoprotein. Because of the multiple amino acid substitutions which are known to be present in α_1-acid glycoprotein (Emura *et al.*, 1971), the sum of the molar residues of each amino acid of CNBr-III and of CNBr-IV is not the same as the number of the corresponding residues of CNBr-II, *e.g.*, CNBr-III possesses three tyrosine and CNBr-IV one, while CNBr-II possesses only three of these residues. For the same reason and because the amino acid compositions of proteins are usually presented as integers, the composition of α_1-acid glycoprotein reported earlier by several investigators (Bourrillon and Meyer, 1967; Marshall and Porath, 1965; Schmid *et al.*, 1968) differed significantly from each other.

Fractionation of the CNBr Fragments after Reduction and Alkylation of the CNBr-Treated Glycoprotein. The CNBr-treated, reduced and carboxymethylated glycoprotein yielded on gel filtration four fractions (I, II, III, and IV) (Figure 1). A fifth fraction (not indicated in Figure 1) contained not only the excess of the reagents but also the cleaved NANA. The estimated molecular weights, recovery, amino- and carboxyl-terminal amino acids, and carbohydrate composition of these fragments are given in Table I. It is important to note that fragment I carries the total carbohydrate moiety of the protein. Therefore, the carbohydrate content of this fragment is higher than that of the native protein. The amino acid compositions of the four fragments are given in Table II.

Fragment I, like the original protein, was characterized by the absence of a free amino-terminal amino acid indicating that this fragment is indeed derived from the amino terminus

TABLE II: Amino Acid Composition of α_1-Acid Glycoprotein (AG), the CNBr-Treated Glycoprotein (CNBr-AG) and the Resulting Four Fragments (I, II, III, and IV).

	Expressed in Moles/Mole of Compound					
			CNBr Fragments			
Amino Acid	AG[a]	CNBr-AG[b]	I[c]	II	III	IV
CM-Cysteine	0	0	3	2	12	1
Aspartic acid	23	24	15	9	6	2
Threonine	16	16	13	3	2	1
Serine	8	8	6	3	3	1
Glutamic acid	31	32	17	14	6	8
Proline	9	9	6	3	2	1
Glycine	9	9	5	3	3	1
Alanine	11	10	7	3	2	0
Valine	10	10	6	4	1	0
Methionine	1.3[d]	0	0	0	0	0
Isoleucine	10	10	9	1	1	0
Leucine	16	15	10	6	5	1
Tyrosine	12	12	9	3	3	1
Phenylalanine	10	10	8	2	3	0
Lysine	16	17	7	9	3	5
Histidine	4	4	3	1	0	1
Arginine	10	10	8	2	1	1
Tryptophan	4	4	1	2	1	1
Half-cystine	6	(6)[e]	0	0	0	0
Homoserine	0	1.4	1	0	1	0
Total	207	208	134	70	45	25

[a] Calculated for a molecular weight of 40,000. These values represent the average of several published analyses (Bourrillon and Meyer, 1967; Marshall and Porath, 1965; Schmid *et al.*, 1968). [b] Average of two analyses. Although about one-third of the sialic acid residues was cleaved, the molecular weight of the polypeptide moiety remained unchanged. [c] Calculated for a molecular weight of 28,000. [d] Schmid *et al.* (1968). [e] Assumed. Marshall and Porath (1965) reported four half-cysteine residues. [f] See text for conversion of homoserine lactone to homoserine.

of α_1-acid glycoprotein (Ikenaka *et al.*, 1966). Further evidence in support of this assumption was the finding that the other three fragments had free α-amino groups. The presence of 1 mole of homoserine/mole of fragment I suggested a high degree of purity of this preparation and complete cleavage of the corresponding methionyl residue. The high degree of homogeneity was further attested by the mentioned absence of a reactive α-amino group. On ultracentrifugal analysis fragment I revealed a single refractive index gradient which sedimented with a coefficient of 2.8 S at a concentration of 1%. This fragment also appeared homogeneous on paper and starch gel electrophoresis at pH 8.6. The characterization of fragments I, III, and IV will be described later with regard to their amino acid sequences. It should be added that fragment III is insoluble in acid solutions so that it would not be recovered if, for the separation of the CNBr fragments from each other, 5–10% acetic acid was used as solvent.

Fragment II appeared homogeneous at pH 8.6 on paper and starch gel electrophoresis in 8 M urea. The electrophoretic mobility of this fragment was higher than that of the CNBr-

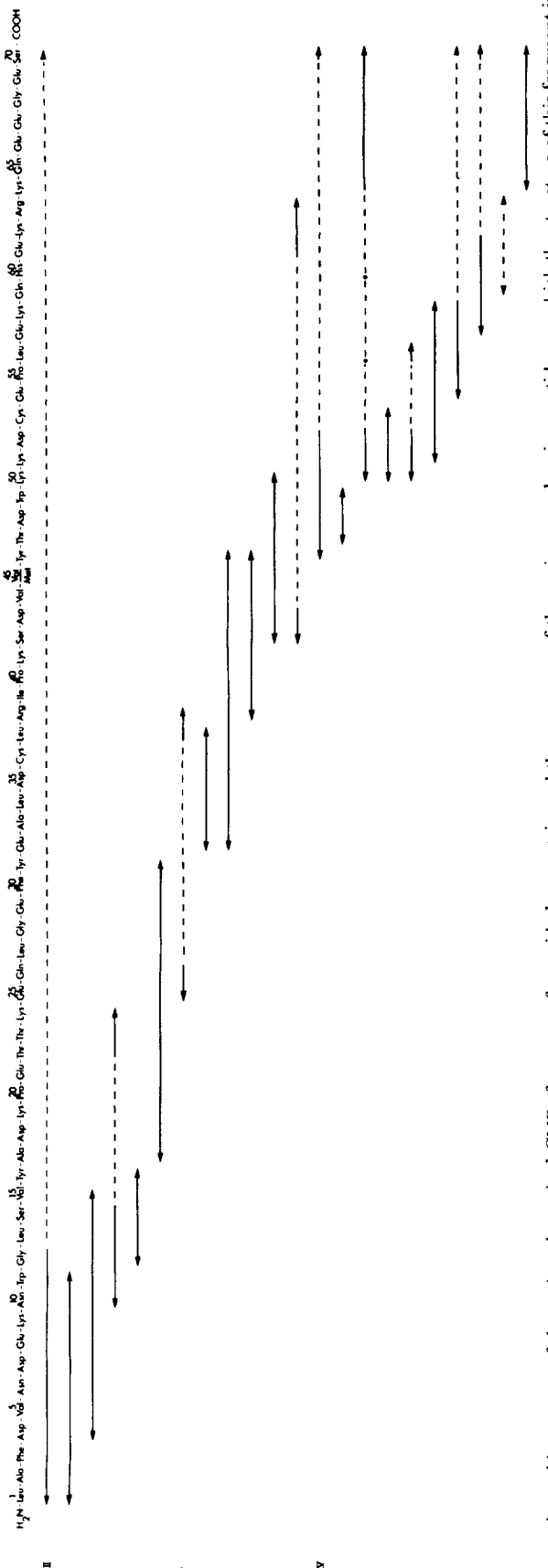

FIGURE 2: The amino acid sequence of the carboxyl-terminal CNBr fragment of α_1-acid glycoprotein and the sequences of the various overlapping peptides on which the structure of this fragment is based. The solid lines represent elucidated sequences. The broken lines indicate that only the amino acid compositions of part of some of their peptides have been established. (▶) Carboxyl-terminal serine was not determined. (—●—) (C3) sequence of the peptides was established.

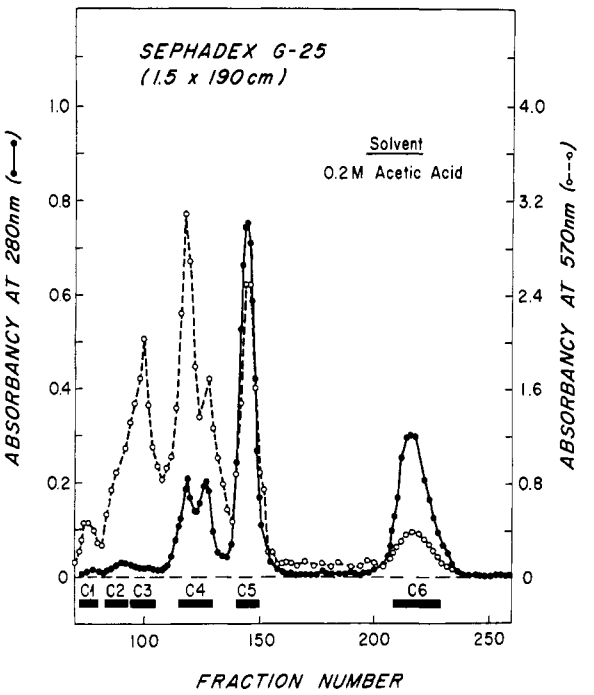

FIGURE 3: Fractionation of a chymotryptic digest of fragment II by gel filtration. Fractions of 2 ml of the effluent were collected, and the flow rate was 7 ml/hr. The effluent was monitored at 280 nm with an LKB Uvicord and by the ninhydrin reaction after alkaline hydrolysis (Yemm and Cocking, 1967; Hirs, 1967) of appropriate aliquots of the effluent.

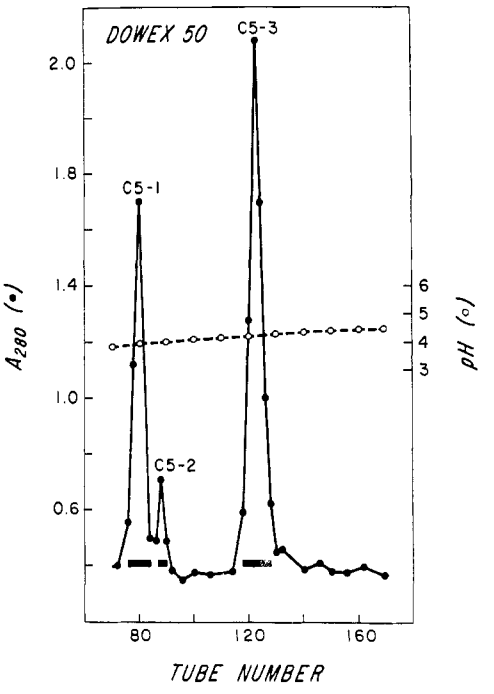

FIGURE 4: Chromatographic separation of the three peptides of fraction C5 (Figure 1) on a Bio-Rad AG 50W-X2 column. The elution was carried out at 38°. Fraction C5 (7 mg) yielded 2.3 mg of C5-1, 0.7 mg of C5-2, and 3.0 mg of C5-3.

treated glycoprotein, compensating for the lower mobility of fraction I. On starch gel electrophoresis at pH 3.3 in 8 M urea homogeneity was also observed. The homogeneity of this fragment was further demonstrated by the sole amino-

FIGURE 5: Homology between the segment of residues 1–21 of the carboxyl-terminal CNBr fragment of α-acid glycoprotein (AG) and a segment of the α chain of haptoglobin (Hapt). For this presentation the single-letter code for amino acids is used (Dayhoff, 1969).

TABLE III: Amino Acid Composition[a] of Chymotryptic Peptides of CNBr Fragment II of α_1-Acid Glycoprotein.

Amino Acid	Residues/Molecule							Sum of Composition of Chymotryptic Peptides	Composition of Fragment II
	C5-3	C5-1	C4-1	C4-4	C4-3	C6	C3		
CM-Cysteine				0.87 (1)			0.90 (1)	2	1.81 (2)
Aspartic acid	3.96 (4)		1.15 (1)	1.00 (1)	1.04 (1)	1.00 (1)	1.20 (1)	9	9.33 (9)
Threonine			1.54[a] (2)			0.96 (1)		3	3.08 (3)[a]
Serine		1.00 (1)			0.72[a] (1)		1.25 (1)	3	2.80 (3)[a]
Glutamic acid	1.07 (1)		3.90 (4)	1.00 (1)			8.10 (8)	14	14.90 (14)
Proline			1.20 (1)		1.12 (1)		1.09 (1)	3	3.31 (3)
Glycine		1.01 (1)	1.00 (1)				1.15 (1)	3	2.98 (3)
Alanine	1.00 (1)		1.10 (1)	1.00 (1)				3	3.08 (3)
Valine	1.00 (1)	0.98 (1)			1.85[a] (2)			4	3.21 (4)[a]
Isoleucine					1.00 (1)			1	1.10 (1)
Leucine	0.95 (1)	1.00 (1)	0.90 (1)	1.90 (2)			1.07 (1)	6	5.47 (6)[a]
Tyrosine		1.20 (1)	0.88 (1)		0.94 (1)			3	2.71 (3)[a]
Phenylalanine	0.84 (1)		0.81 (1)					2	1.85 (2)
Lysine	1.04 (1)		1.91 (2)		1.03 (1)		4.85 (5)	9	9.00 (9)
Histidine							(1)	1	0.98 (1)
Arginine					0.89 (1)		0.98 (1)	2	1.91 (2)
Tryptophan	0.20[a] (1)					0.30[a] (1)		2	1.80 (2)[b]
Total number of residues	11	5	15	6	9	3	21	70	70
Yield (%)	40	55	48	50	60	40	30		
Position in sequence	1–11	12–16	17–31	32–27	38–46	47–49	50–70		

[a] All amino acid analyses were carried out on 24-hr hydrolysates. No corrections were applied for partial destruction of threonine, serine and tyrosine and for partial hydrolysis of valine and leucine. The tryptophan content of carbohydrate-free peptides may be recovered between 20 and 30% (Bradshaw *et al.*, 1969). [b] Spectrophotometrically according to Bencze and Schmid (1957).

TABLE IV: Amino Acid Sequence of Peptide C5-3-C2.

Sequence	Asp-Val-Asn-Asp-Glu-Lys-Asn-Trp
CPA	2 hr: Trp, 0.95; Asn, 0.30
Edman degradation[a]	
Step 1	*Asp*, 1.10; Val, 0.95; Asn, 2.10; Glu, 0.95; Lys, 0.65
Step 2	Asp, 0.95; *Val*, 0.12; Asn, 2.10; Glu, 0.95; Lys, 0.46
Step 3	Asp, 0.90; *Asn*, 1.05; Glu, 1.00; Lys, 0.36
Step 4	*Asp*, 0.20; Asn, 0.95; Glu, 0.97; Lys, 0.30
Step 5	Asn, 0.95; *Glu*, 0.20; Lys, 0.20
Step 6	Asn, 0.97; *Lys*, 0.00

[a] The peptides obtained after each Edman step were completely digested with AP-M. The obtained amino acids were separated from each other by pc in BAW. Tryptophan was not determined.

TABLE V: Amino Acid Sequence of Peptide C4-1-T1.[a]

Sequence	Glu-Gln-Leu-Gly-Glu-Phe-Tyr
Edman degradation	
Step 1	*Glu*, 1.35; Gln, 0.83; Leu, 0.87; Gly, 0.95; Phe, 1.00; Tyr, 1.17
Step 2	Glu, 1.35; *Gln*, 0.01; Leu, 1.00; Gly, 1.01; Phe, 1.00; Tyr, 1.11
Step 3	Glu, 1.27; *Leu*, 0.28; Gly, 1.07; Phe, 1.00; Tyr, 1.14
Step 4	Glu, 1.30; *Gly*, 0.30; Phe, 1.00; Tyr, 0.97
Step 5	*Glu*, 0.19; Phe, 1.00; Tyr, 0.95

[a] An aliquot of the peptide obtained after each step was digested completely with AP-M (24 hr, 37°) and subjected to paper chromatography in BAW in order to establish the position of the amides.

and carboxyl-terminal amino acids (Table I) and by the fact that in 13 consecutive Edman cycles, each yielded a single amino acid (see below).

Elucidation of the Amino Acid Sequence of Fraction II

Fragment II. Fragment II (0.5 μmole), when subjected to the direct Edman procedure, proved to possess the following partial amino-terminal sequence

$$\underset{0.95}{\overrightarrow{\text{Leu}}}\text{-}\underset{0.95}{\overrightarrow{\text{Ala}}}\text{-}\underset{0.75}{\overrightarrow{\text{Phe}}}\text{-}\underset{0.63}{\overrightarrow{\text{Asp}}}\text{-}\underset{0.48}{\overrightarrow{\text{Val}}}\text{-}\underset{0.46}{\overrightarrow{\text{Asn}}}\text{-}\underset{0.46}{\overrightarrow{\text{Asp}}}\text{-}\underset{0.33}{\overrightarrow{\text{Glu}}}\text{-}\underset{0.24}{\overrightarrow{\text{Lys}}}\text{-}\underset{0.12}{\overrightarrow{\text{Asn}}}\text{-}\underset{0.15}{\overrightarrow{\text{Trp}}}\text{-}\underset{0.12}{\overrightarrow{\text{Gly}}}\text{-}\underset{0.09}{\overrightarrow{\text{Leu}}}$$

This and all sequences established subsequently are indicated in Figure 2.

Chymotryptic Peptides. The chymotryptic digest of fragment II was resolved by gel filtration into six fractions (Figure 3). On hve at pH 1.9 and 6.4 all fractions proved to be mixtures except two, namely, fractions C3 and C6. The three peptides of fraction C4 were isolated in pure form by hve at 6.4. The peptides of fraction C5 were purified by ion-exchange chromatography on Dowex 50W-X2 (Figure 4). The amino acid compositions and recoveries of these peptides are listed in Table III.

PEPTIDE C5-3 (Residues 1–11). Four amino-terminal residues established by the direct Edman procedure and two carboxyl-terminal residues established by CPA digestion for 2 hr showed that this peptide forms the amino terminus of fragment II

$$\underset{0.78}{\overrightarrow{\text{Leu}}}\text{-}\underset{0.54}{\overrightarrow{\text{Ala}}}\text{-}\underset{0.19}{\overrightarrow{\text{Phe}}}\text{-}\underset{0.16}{\overrightarrow{\text{Asp}}}\text{-}\underset{0.04}{\overrightarrow{\text{Val}}}\text{-(Asn,Asp,Glu,Lys)-}\underset{0.52}{\overleftarrow{\text{Asn}}}\text{-}\underset{0.97}{\overleftarrow{\text{Trp}}}$$

PEPTIDE C5-3-C2 (Residues 4–11; Table VI). The sequence of this Ehrlich-positive, acidic peptide derived from a chymotryptic digest of peptide C5-3 was elucidated by six successful subtractive Edman steps (Table IV).

Peptide C5-1 (Residues 12–16). The sequence of this peptide was established by two direct Edman steps and by CPA digestion (Figure 2).

$$\underset{0.44}{\overrightarrow{\text{Gly}}}\text{-}\underset{0.21}{\overrightarrow{\text{Leu}}}\text{-}\underset{0.25}{\overrightarrow{\text{Ser}}}\text{-}\underset{0.70}{\overleftarrow{\text{Val}}}\text{-}\underset{1.00}{\overleftarrow{\text{Tyr}}}\text{-}$$

This structure was confirmed by five successful subtractive Edman steps.

PEPTIDE C4-1 (Residues 17–31). The partial sequence of this peptide was established by seven direct Edman steps and by analysis of a CPA digest withdrawn after 5-, 30-, and 120-min incubation.

$$\underset{1.00}{\overrightarrow{\text{Ala}}}\text{-}\underset{0.75}{\overrightarrow{\text{Asp}}}\text{-}\underset{0.71}{\overrightarrow{\text{Lys}}}\text{-}\underset{0.50}{\overrightarrow{\text{Pro}}}\text{-}\underset{0.44}{\overrightarrow{\text{Glu}}}\text{-}\underset{(*)}{\overrightarrow{\text{Thr}}}\text{-}\underset{(*)}{\overrightarrow{\text{Thr}}}\text{-(Lys,Glu,Gln,Leu,Gly,Glu)-}\underset{0.90}{\overleftarrow{\text{Phe}}}\text{-}\underset{0.95}{\overleftarrow{\text{Tyr}}}$$

Following tryptic digestion of C4-1 and purification of the formed peptides, the carboxyl-terminal peptide (C4-1-T2) was elucidated by the subtractive Edman degradation method (Table V). To complete the sequence of C4-1, an aliquot of this peptide obtained after each Edman step was digested completely with AMP (24 hr, 37°) and subjected to paper chromatography in BAW in order to establish the position of the amides. Confirmatory evidence of the structure of C4-1 was obtained from the amino acid composition of the three peptides resulting from a thermolysin digest, including peptide C4-1-Th2.

PEPTIDE C4-4 (Residues 32–37). Four steps of direct Edman degradation and digestion with CPA for 4 hr established the sequence as follows

$$\underset{0.91}{\overrightarrow{\text{Glu}}}\text{-}\underset{0.82}{\overrightarrow{\text{Ala}}}\text{-}\underset{0.46}{\overrightarrow{\text{Leu}}}\text{-}\underset{0.49}{\overrightarrow{\text{Asp}}}\text{-Cys-}\underset{0.62}{\overleftarrow{\text{Leu}}}$$

The position of CM-cysteine was thus unequivocally determined.

PEPTIDE C4-3 (Residues 38–46). Six steps of direct Edman degradation and CPA digestion established the sequence of this peptide as follows

$$\underset{(+)}{\overrightarrow{\text{Arg}}}\text{-}\underset{0.69}{\overrightarrow{\text{Ile}}}\text{-}\underset{0.43}{\overrightarrow{\text{Pro}}}\text{-}\underset{0.32}{\overrightarrow{\text{Lys}}}\text{-}\underset{(*)}{\overrightarrow{\text{Ser}}}\text{-}\underset{0.20}{\overrightarrow{\text{Asp}}}\text{-}\underset{0.50}{\overleftarrow{\text{Val}}}\text{-}\underset{1.00}{\overleftarrow{\text{Val}}}\text{-}\underset{1.00}{\overleftarrow{\text{Tyr}}}$$

From an additional, limited chymotryptic digest of frag-

TABLE VI: Amino Acid Composition[a] of Certain Additional Chymotryptic Peptides and a Tryptic Peptide (T1) of CNBr-Fragment II.

Amino Acid	Residues/Molecule				
	CA	CC	CB	C5-3-C2	T1
Cm-Cysteine	0.83 (1)	0.82 (1)	0.60 (1)		
Aspartic acid	2.83 (3)	2.10 (2)	2.04 (2)	4.12 (4)	2.00 (2)
Threonine[a]	1.63 (2)	1.80 (2)			1.00 (1)
Serine[a]	0.81 (1)		0.89 (1)		1.03 (1)
Glutamic acid	4.65 (5)	4.82 (5)	0.91 (1)	0.91 (1)	
Proline	1.83 (2)	1.20 (1)	1.00 (1)		
Glycine	1.00 (1)	1.00 (1)			
Alanine	1.86 (2)	1.92 (2)	1.00 (1)		
Valine[a]	1.10[b] (2)		1.80 (2)	1.02 (1)	1.23 (2)
Isoleucine[a]	0.74 (1)		0.91 (1)		
Leucine[a]	2.46 (3)	2.84 (3)	2.04 (2)		
Tyrosine[a]	1.27 (2)	0.82 (1)	0.94 (1)		0.97 (1)
Phenylalanine	0.89 (1)	0.91 (1)			
Lysine	2.67 (3)	2.03 (2)	1.25 (1)	1.03 (1)	1.06 (1)
Histidine					
Arginine	0.82 (1)		1.08 (1)		
Tryptophan				0.25 (1)	
Total number of residues	30	21	15	8	8
Yield (%)	7	8	11	10	15
Position in Sequence	17–46	17–37	32–46	4–11	42–50

[a] See Table III, footnote *a*. [b] A Val–Val bond is known to yield only about 60% free valine during a 24-hr hydrolysis.

ment II a peptide designated CB (residues 32–46; Table VI) was isolated which had the composition of C4-4 plus C4-3. The partial sequence of this peptide elucidated by 7 subtractive Edman degradation steps (Glu-Ala-Leu-Asp-Cys-Leu-Arg-) and by digestion with CPA (-Val-Tyr) establishing the amino and carboxyl terminus of this peptide. From a tryptic digest of CB two peptides were isolated. Their amino acid composition agreed with the two sequences: from residues 39–41 and 42–46. From a thermolytic digest of CB four peptides were isolated, and their composition agreed with the sequence from residues 32–33, 34–36, 37–43, and 44–46, respectively. From these data, including a consideration of the specificity of the enzymes employed, the structure of CB was deduced as indicated in Figure 2. This sequence confirmed those of C4-4 and C4-3 and indicated their positions in fragment II. An additional, larger peptide (CA, residues 17–46; Table VI), accounted for the composition of the peptides C4-1 plus C4-4 plus C4-3. Five residues established by the direct Edman procedure

$$\underset{0.90}{\text{Ala}} - \underset{0.81}{\text{Asp}} - \underset{0.73}{\text{Lys}} - \underset{0.65}{\text{Pro}} - \underset{0.59}{\text{Glu}}-$$

were identical with those of C4-1 and, hence, indicated the following peptide sequence: (C4-1)-(C4-4)-(C4-3). A further peptide, CC (residues 17–37), had the composition of C4-1 plus C4-4 (Table VI) and possessed the same four amino-terminal amino acids (direct Edman) as peptide C4-1 confirming this peptide sequence. An even more important overlapping peptide, LC4-T2, is discussed below.

PEPTIDE C6 (Residues 47–49). Two direct Edman steps and CPA digestion for 2 hr of this Ehrlich-positive peptide established the sequence as follows

$$\underset{(*)}{\text{Thr}} - \underset{\substack{0.40 \\ 0.80\ \ 0.95}}{\text{Asp}} - \text{Trp}$$

PEPTIDE C3 (Residues 50–70). The partial sequence of this peptide was established as follows. Direct Edman degradation yielded the sequence of three residues

$$\underset{0.80}{\text{Lys}} - \underset{0.60}{\text{Lys}} - \underset{0.65}{\text{Asp}}$$

It is noteworthy that the carboxyl-terminal serine of this peptide as well as that of the whole CNBr fragment and α_1-acid glycoprotein was not cleaved by CPA. However, on prolonged digestion with CPC (17 hr) a minute amount of serine and glutamic acid was released. In an earlier study (Schmid *et al.*, 1959), the carboxyl-terminal serine of this glycoprotein was established by an independent procedure.

Peptide C3 was subsequently digested with trypsin. One peptide (peptide C3–T3; residues 64–70) was isolated and its sequence established by six direct Edman steps

$$\underset{0.66}{\text{Lys}} - \underset{0.50}{\text{Gln}} - \underset{0.24}{\text{Glu}} - \underset{0.22}{\text{Glu}} - \underset{0.10}{\text{Gly}} - \underset{0.03}{\text{Glu}} - \text{Ser}$$

(See below peptide LC6-b, especially for determination of the carboxyl-terminal serine.) From a thermolytic digest of C3 two peptides (C3-Th-1 and C3-Th-2) were isolated and their amino acid compositions was determined (Table VII) and found

TABLE VII: Amino Acid Composition[a] of Chymotryptic (C) and Thermolytic (Th) Peptides of Peptide C3.

Amino Acid	Residues/Molecule					
	CIII-a	CIII-b	CIII-c	CIII-d	C3-Th-1	C3-Th-2
CM-Cysteine	0.60 (1)	+ (1)			+ (1)	
Aspartic acid	1.00 (1)	1.01 (1)			0.97 (1)	
Threonine						
Serine			0.76 (1)	0.89 (1)		
Glutamic acid		0.96 (1)	9.7 (8)	9.85 (7)	1.0 (1)	1.67 (2)
Proline		1.00 (1)	1.38 (1)		1.63 (1)	
Glycine			0.85 (1)	0.87 (1)		
Alanine						
Valine						
Isoleucine						
Leucine		0.93 (1)	0.93 (1)			1.00 (1)
Tyrosine						
Phenylalanine						
Lysine	2.18 (2)	1.95 (2)	3.46 (3)	3.03 (3)	2.05 (2)	0.75 (1)
Histidine			0.54 (1)	1.18 (1)		
Arginine			1.00 (1)	1.00		
Tryptophan						
Total number of residues	4	7	17	14	6	4
Yield	8	3	8	7		
Position in sequence	50–53	50–56	54–70	57–70	50–55	56–59

[a] See footnote *a* of Table III.

to be in agreement with the sequences of the residues from 50 to 55 and 56 to 59, respectively.

From a third chymotryptic digest of fragment II four peptides were isolated by chromatography on Sephadex G-25 and CM-cellulose at pH 4.0 and 5.0. These compounds, which were designated as CIII-a, CIII-b, CIII-c and CIII-d and whose amino acid compositions are listed in Table VII, distinguished themselves by the lack of absorbancy at 280 nm (like fraction C3; Figure 2).

PEPTIDE CIII-a (Residues 50–53). This peptide appeared neutral when analyzed on hve at pH 6.4, and, hence, does not possess any asparagine. Three successful direct Edman steps afforded the following sequence

$$\underset{0.57}{\overrightarrow{\text{Lys}}}\text{ - }\underset{0.36}{\overrightarrow{\text{Lys}}}\text{ - }\underset{0.38}{\overrightarrow{\text{Asp}}}\text{ - Cys-}$$

Consequently, the CM-cysteine residue must form the carboxyl terminus. Digestion with CPA did not release any amino acids.

PEPTIDE CIII-b (Residues 50–56). Partial elucidation of the structure of this peptide was achieved by three successful direct Edman steps and digestion with CPA and CPC

$$\underset{0.75}{\overrightarrow{\text{Lys}}}\text{ - }\underset{0.53}{\overrightarrow{\text{Lys}}}\text{ - }\underset{0.31}{\overrightarrow{\text{Asp}}}\text{ (Cys,Glu)}\underset{0.27}{\overleftarrow{\text{Pro}}}\text{ - }\underset{0.72}{\overleftarrow{\text{Leu}}}$$

Carboxypeptidase A yielded a minute but distinct amount of leucine. However, CPC released leucine (0.72) and proline (0.27).

PEPTIDE CIII-c (Residues 54–70). Four successful Edman steps elucidated the amino terminus of this peptide

$$\underset{0.90}{\overrightarrow{\text{Glu}}}\text{ - }\underset{0.67}{\overrightarrow{\text{Pro}}}\text{ - }\underset{0.46}{\overrightarrow{\text{Leu}}}\text{ - }\underset{0.48}{\overrightarrow{\text{Glu}}}\text{ - }\underset{0.32}{\overrightarrow{\text{Lys}}}\text{-}$$

PEPTIDES CIII-d (Residues 57–70). The partial sequence of this peptide was obtained as follows. First, direct Edman degradation allowed establishing the sequence of three residues

$$\underset{0.95}{\overrightarrow{\text{Glu}}}\text{ - }\underset{0.67}{\overrightarrow{\text{Lys}}}\text{ - }\underset{0.15}{\overrightarrow{\text{Gln}}}\text{-}$$

The peptide obtained after three Edman steps was subsequently digested with AP-M. Analysis of aliquots of the resulting digest withdrawn after different incubation times established the sequence of the next two residues: His-Glu. The carboxyl-terminal amino acid of this peptide was established by a successful Stark degradation step. The cleaved serine derivative afforded the expected two spots on tlc (Cromwell and Stark, 1969). In order to confirm the removal of the carboxyl-terminal serine by this technique, the degraded peptide was analyzed for its amino acid composition. The content of serine was reduced to approximately half, while that of the other amino acids remained unchanged.

Tryptic Peptides of Citraconylated Fragment II. Gel filtration of a tryptic digest (100 mg) of the citraconylated fragment yielded three major fractions (LC4, 14 mg; LC5, 34 mg; LC6, 12 mg) of which LC4 proved to be homogeneous as judged by hve at pH 6.4. An aliquot of peptide LC4 was digested with trypsin. Isolation of the resulting two major peptides (LC4-T1 and LC4-T2) was achieved by gel filtration through Sephadex G-25 in 0.5% ammonium bicarbonate buffer (pH 8.0). While peptide LC4-T1 was obtained in homogeneous form, peptide LC4-T2 was obtained as a monodisperse compound after purification by hve at 3.5. The two major peptides of fraction LC5 were separated from each other by chromatography on a DEAE-cellulose column. Peptide LC5-1 could be purified by hve at pH 6.4 and peptide

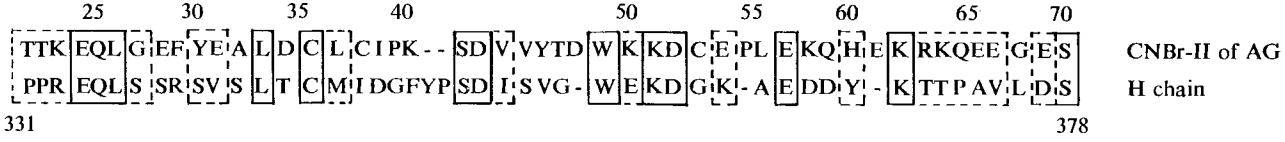

FIGURE 6: Homology between segment of residues 22–70 of the carboxyl-terminal CNBr fragment of α_1-acid glycoprotein (AG), and the constant region of the H chain of immunoglobulin. Identical residues are boxed in. Residues, that would be identical if a single point mutation had occurred, are in boxes with broken frames.

TABLE VIII: Amino Acid Composition of Tryptic Peptides of the Citraconylated CNBr Fragment II.

	Residues/Molecule							
Amino Acid	LC4	LC4-T1	LC4-T2	LC6-a	LCG-c	LC5-2	LC5-1	LC6-b
CM-Cysteine	1.10 (1)		0.82 (1)			1.05 (1)		
Aspartic acid	5.32 (6)	1.90 (2)	1.26 (1)		2.21 (2)	3.03 (3)		
Threonine[a]	1.83 (2)	1.95 (2)			0.95 (1)	1.07 (1)		
Serine[a]	1.15 (1)	0.78 (1)			0.91 (1)	0.88 (1)		0.84 (1)
Glutamic acid	6.24 (6)	2.00 (1)	3.88 (4)			4.30 (4)	1.90 (2)	3.96 (4)
Proline	0.90 (1)	1.25 (1)		1.12 (1)		1.25 (1)		
Glycine	1.76 (2)	0.92 (1)	1.25 (1)					1.3 (1)
Alanine	2.62 (3)	1.10 (1)	1.24 (1)					
Valine[a]	1.80 (2)	1.13 (1)			2.09 (2)	1.94 (2)		
Isoleucine[a]				1.00 (1)				
Leucine[a]	4.36 (5)	1.00 (1)	3.02 (3)			1.63 (1)		
Tyrosine[a]	1.95 (2)	0.83 (1)	0.86 (1)		0.83 (1)	1.20 (1)		
Phenylalanine	1.76 (2)		1.35 (1)					
Lysine	2.84 (3)	2.30 (2)		1.00 (1)	2.00 (2)	5.30 (5)	1.00 (1)	1.00 (1)
Histidine						1.00 (1)	++ (1)	
Arginine	1.00 (1)		+ (1)			1.00 (1)	+ (1)	
Tryptophan	+++ (1)	ND (1)				ND (1)		
Total number of residues	38	15	14	3	9	23	5	7
Yield (%)	64	20	10	20	12	15	9	30
Position in sequence	1–38	10–24	25–38	39–41	42–51	42–63	59–63	64–70

[a] See Table III, footnote *a*; ND, not determined; +, Sackaguchi positive; ++, Pauli positive; +++, Ehrlich positive.

TABLE IX: Amino Acid Sequence of Peptide P1.

Composition	Asx	Val	Glx	Lys	Trp	Gly	Leu	Ser
Sequence	Asp-Val-Asn-Asp-Glu-Lys-Asn-Trp-Gly-Leu-Ser-Val							
Edman degradation								
Step 0	4.14	2.38	1.00	1.00	ND	0.98	1.00	0.78
Step 1	*2.95*	1.92	0.79	0.64	ND	1.00	1.00	0.76
Step 2	2.87	*1.21*	0.78	0.52	ND	0.82	1.00	0.72
Step 3	*1.85*	1.19	0.66	0.22	ND	0.86	1.00	0.78
Step 4	*1.20*	1.18	0.87	0.25	ND	1.08	1.00	0.81
Step 5	1.12	1.00	*0.18*	0.36	ND	1.01	1.00	0.67
Step 6	0.84	1.13		*0.00*	ND	0.97	1.00	0.80
Step 7	*0.19*	1.13			ND	1.06	1.00	0.71
Step 8		1.00			*ND*[a]	1.18	1.00	0.78
Step 9		1.26			ND	*0.05*	1.00	0.78
Step 10		1.00			ND		*0.11*	0.73
Step 11		1.00			ND			*0.29*

[a] The composition of the peptide resulting from step 8 was essentially identical with that obtained in the previous step. Hence, it was assumed that tryptophan was removed at step 8.

LC5-2 by gel filtration through a Sephadex G-25 column. The three major peptides of fraction LC6 (LC6-a, LC6-b, and LC6-c) were isolated by hve at pH 6.4 and 1.8. From a further fraction peptide LC7 was isolated. The partial sequences of these peptides whose amino acid compositions are listed in Table VIII were established as follows.

PEPTIDE LC4 (Residues 1–38). Two steps of direct Edman degradation

$$\underset{0.55}{\overrightarrow{\text{Leu}}}\text{ - }\underset{0.47}{\overrightarrow{\text{Ala}}}$$

digestion with CPB for 2 hr followed with CPA for 17 hr,

$$(\text{-}\underset{0.86}{\overleftarrow{\text{Leu}}}\text{-}\underset{1.00}{\overleftarrow{\text{Arg}}})$$

and the amino acid composition demonstrated that this peptide was derived from the amino terminus of the original fragment.

PEPTIDE LC4-T1 (Residues 10–24). Five successful subtractive Edman steps and digestion with CPB followed with CPA established the amino and carboxyl termini of this peptide and its position within the sequence of CNBr fragment II.

PEPTIDE LC4-T2 (Residue 25–38). Two direct Edman steps

$$(\underset{0.70}{\overrightarrow{\text{Glu}}}\text{ - }\underset{0.40}{\overrightarrow{\text{Gln}}})$$

and digestion with CPB (2 hr) and CPA (17 hr)

$$(\underset{0.85}{\overleftarrow{\text{Leu}}}\text{ - }\underset{1.00}{\overleftarrow{\text{Arg}}})$$

established both termini of this peptide and, thus, the position within the sequence of fragment II.

PEPTIDE LC6-a (Residue 39–41). Because of the specificity of trypsin it was assumed that lysine formed the carboxyl terminus of this peptide. The sequence of this composition was established by three subtractive Edman steps.

PEPTIDE LC6-e (Residues 42–51). The amino acid composition of this peptide was in agreement with the structure of the corresponding segment of fragment II.

PEPTIDE LC5-2 (Residues 42–63). Two successful direct Edman steps

$$(\underset{(*)}{\overrightarrow{\text{Ser}}}\text{ - }\underset{0.10}{\overrightarrow{\text{Asp}}}\text{-})$$

and digestion with CPB for 3 hr followed with CPA for 4 hr

$$(\underset{0.67}{\overleftarrow{\text{Glu}}}\text{ - }\underset{0.91}{\overleftarrow{\text{Lys}}}\text{ - }\underset{1.00}{\overleftarrow{\text{Arg}}})$$

allowed unequivocal positioning of this peptide within the sequence of fragment II (Figure 2).

PEPTIDE LC7 (Residues 39–63). Because of partial substitution of the ϵ-amino group of lysine-41, a peptide with the composition of peptides LC6-a plus LC5-Cl could also be isolated.

PEPTIDE LC5-1 (Residue 59–63). The amino acid composition of this peptide was found to be in agreement with the corresponding sequence of fragment II.

PEPTIDE LC6-b (Residues 64–70). The sequence of this peptide[6] was established by the direct Edman technique

$$\underset{0.66}{\overrightarrow{\text{Lys}}}\text{ - }\underset{0.75}{\overrightarrow{\text{Gln}}}\text{ - }\underset{0.48}{\overrightarrow{\text{Glu}}}\text{ - }\underset{0.44}{\overrightarrow{\text{Glu}}}\text{- }\underset{0.20}{\overrightarrow{\text{Gly}}}\text{ - }\underset{0.09}{\overrightarrow{\text{Glu}}}\text{ - }\underset{(*)}{\overrightarrow{\text{Ser}}}$$

The carboxyl-terminal amino acid was established by the Stark procedure. The results obtained were identical with those described above for peptide CIII-d.

Peptic Peptides. A small number of peptides were isolated from a peptic digest of fragment II. The peptide designated P1 was of particular interest as it confirmed the sequence from the residues 4–15. This Ehrlich-positive, acidic compound was isolated in a yield of 40%. Since it was free of arginine and histidine, the peptides resulting from the subtractive Edman steps could be carried out by glpc. Eleven successful subtractive Edman steps (Table IX) and digestion

[6] This sequence was earlier established by Collins (1970).

with CPA yielded the sequence of this peptide. Direct Edman degradation established residue 1 to be Asp and residue 3 to be Asn. The sequence of this peptide is therefore as follows: Asp-Val-Asn-Asp-Glu-Lys-Asn-Trp-Gly-Leu-Ser-Val.

Tryptic Peptides. During the earlier phase of this study, several peptides were isolated from a tryptic digest of fragment II. Of particular interest was peptide T1 as it represents an important overlapping peptide whose structure is required to ascertain the sequence in the region of residue number 46.

PEPTIDE T1 (Residues 42–50). The structure of this Ehrlich-positive peptide was established by eight direct Edman steps. Because of the specificity of trypsin, it was assumed that the sole lysine residue represents the carboxyl terminus. Therefore, tryptophan must be the penultimate residue. The sequence of this peptide is therefore as follows

$$\underset{(*)\ 0.48}{\overrightarrow{\text{Ser}}}\text{-}\underset{0.42}{\overrightarrow{\text{Asp}}}\text{-}\underset{0.29}{\overrightarrow{\text{Val}}}\text{-}\underset{0.29}{\overrightarrow{\text{Val}}}\text{-}\underset{0.30}{\overrightarrow{\text{Tyr}}}\text{-}\underset{0.15}{\overrightarrow{\text{Thr}}}\text{-}\underset{0.11}{\overrightarrow{\text{Asp}}}\text{-}\text{Trp-Lys}$$

PEPTIDE T2 (Residues 52–58). Five successful Edman steps and digestion with CPB followed by CPA which released lysine and glutamic acid, respectively, established the structure of this peptide as indicated in Figure 2.

Discussion

The structure of the carboxyl-terminal CNBr fragment of α_1-acid glycoprotein consisting of 70 amino acid residues is characterized by (a) a large number of basic [9 lysines, 1 histidine, and 2 arginines] and acid [7 aspartic and 11 glutamic acid] residues. Near neutrality 30 of the 70 residues are dissociated so that this CNBr fragment becomes highly polar. Moreover, a particularly large number of these residues was found to be concentrated in the carboxyl-terminal 21-residue segment of this fragment. In contrast, the amino-terminal two-thirds of the polypeptide chain of this glycoprotein possesses a relatively small number of dissociable amino acid residues (Schmid *et al.*, 1971).

For the elucidation of the amino acid sequence of this CNBr fragment, trypsin, because of the above-mentioned concentration of lysine residues, especially the two regions Lys-Lys and Lys-Arg-Lys, yielded a peptide mixture and free arginine (Bourrillon *et al.*, 1970) which proved relatively difficult to be elucidated.

Chymotrypsin, however, because of the almost regular distribution of the aromatic amino acids, afforded peptides from which the largest part of the structure could be deduced. An exception was the carboxyl-terminal 21-residue peptide, whose sequence could in part be established from tryptic and thermolytic peptides. It should be noted that this large peptide, being free of tyrosine and tryptophan, does not exhibit any absorbancy at 280 nm, whereby its isolation from chymotryptic digests of this CNBr fragment was considerably facilitated.

For the preparation of overlapping peptides required to establish the total sequence, this CNBr fragment was citraconylated and then digested with trypsin. As the ϵ-amino groups of the lysine residue 41 and 58 were incompletely substituted, additional tryptic peptides with carboxyl-terminal lysine were also obtained.

The most striking observation of this study is the homology of fragment II with haptoglobin and particularly with the immunoglobulins. Thus, α_1-acid glycoprotein is the first single-chain protein that possesses sequence similarities with two other plasma proteins. The amino-terminal 22-residue segment of this fragment has a direct homology of 36% with a certain section of the α chain of haptoglobin (Figure 5). The remaining 48 residues of fragment II possesses 13 residues that are identical with those of a certain section of the H chain of the immunoglobulins (Dayhoff, 1969) (Figure 6). Residues, such as cysteine, tryptophan, lysine, leucine, and aspartic acid that are known to have the lowest degree of mutability (Dayhoff, 1969), have been preserved. If single-point mutations are included in this consideration, an accepted procedure in comparative studies on immunoglobulins, the homology between these two segments increases to 75%. These findings would seem to suggest that α_1-acid glycoprotein probably evolved at least in part from the ancestal immunoglobulin.

Acknowledgments

The authors are very grateful to Professor Y. Matsushima and Mr. T. Koide, Osaka University, Professor M. Funatsu and Dr. M. Ishiguro, Kyushu University, Dr. W. Ritschard, Hoffmann-La Roche, Basel, Dr. M. Elzinga, Boston Biochemical Research Institute, and Mr. G. Crombie and Dr. C. Franzblau for their kindness in carrying out innumerous amino acid analyses. The authors are greatly indebted to Dr. Birger Blombäck, Karolinska Institut, Stockholm, for his advice on the direct Edman procedure. Cohn fraction VI from which α_1-acid glycoprotein was isolated was generously donated by Mr. L. Larsen, Institute of Laboratories, Massachusetts Department of Health, Jamaica Plain, Mass. The gift of citrus carboxypeptidase from Professor H. Zuber, Federal Institute of Technology, Zürich, is also appreciated. The expert technical assistance of Miss Hedy Hediger and Mr. Collin Munro is gratefully acknowledged.

References

Bencze, W., and Schmid, K. (1957), *Anal. Chem. 29*, 1193.
Bradshaw, R. A., Garner, W. H., and Gurd, F. R. N. (1969), *J. Biol. Chem. 244*, 2152.
Bourrillon, R., and Meyer, D. (1967), *Bull. Soc. Chim. Biol. 49*, 1127.
Bourrillon, R., Meyer, D., and Parnaudeau, M. A. (1970), *Biochim. Biophys. Acta 200*, 378.
Bürgi, W., and Schmid, K. (1961), *J. Biol. Chem. 236*, 1066.
Canfield, R. E., and Anfinsen, C. B. (1963), *Proteins 1*, 311.
Collins, J. H. (1970), Ph.D. Thesis, Boston University, Boston, Mass.
Cromwell, D., and Stark, G. R. (1969), *Biochemistry 8*, 4735.
Dayhoff, M. O. (1969), Atlas of Protein Sequence and Structure, Vol. 4, Silver Spring, Md., National Biomedical Research Foundation.
Dixon, H. B. F., and Perham, R. N. (1968), *Biochem. J. 109*, 312.
Elzinga, M., Lai, C. Y., and Hirs, C. H. W. (1968), *Arch. Biochem. Biophys. 123*, 353.
Emura, J., Ikenaka, T., Collins, J. H., and Schmid, K. (1971), *J. Biol. Chem. 246*, 7821.
Fraenkel-Conrat, H. G., Harris, J. I., and Levy, A. L. (1955), *Methods Biochem. Anal. 2*, 359.
Gehrke, C., and Stalling, D. L. (1970), *Separ. Sci. 11*, 101.
Hirs, W. C. H. (1967), *Methods Enzymol. 11*, 1.
Huang, T. S., and DeLange, R. J. (1971), *J. Biol. Chem. 246*, 686.
Ikenaka, T., Bammerlin, H., Kaufmann, H., and Schmid, K. (1966), *J. Biol. Chem. 241*, 5560.

Ikenaka, T., Ishiguro, M., Isemura, S., Kaufmann, H., Bauer, W., and Schmid, K. (1971), *Biochem. Biophys. Res. Commun. 42*, 1142.
Ishiguro, M., Yamamoto, A., and Schmid, K. (1969), *Biochim. Biophys. Acta 194*, 91.
Iwanaga, S., Wallén, P., Groendahl, N. J., Henschen, A., and Blombäck, B. (1969), *Eur. J. Biochem. 8*, 189.
Jeanloz, R. W. (1966), *in* The Glycoproteins, Gottschalk, A., Ed., New York, N. Y., Elsevier, p 362.
Kaufmann, H., and Schmid, K. (1972), *Experientia 28*, 24.
Landon, M., Melamed, M. D., and Smith, E. L. (1971), *J. Biol. Chem. 246*, 2360.
Marshall, W. E., and Porath, J. (1965), *J. Biol. Chem. 240*, 209.
Moore, S. (1963), *J. Biol. Chem. 238*, 235.
Popenoe, E. A., and Drew, R. M. (1957), *J. Biol. Chem. 228*, 673.
Ryle, A. P., Sanger, F., Smith, L. F., and Kitai, R. (1955), *Biochem. J. 60*, 541.
Schmid, K. (1954), *Biochim. Biophys. Acta 14*, 437.
Schmid, K., Bencze, W. L., Nussbaumer, T., and Wehrmueller, J. O. (1959), *J. Biol. Chem. 234*, 529.
Schmid, K., Ishiguro, M., Emura, J., Isemura, S., Kaufmann, H., and Motoyama, T. (1971), *Biochem. Biophys. Res. Commun. 242*, 280.
Schmid, K., Okuyama, T., and Kaufmann, H. (1968), *Biochim. Biophys. Acta 154*, 565.
Schroeder, W. A. (1967), *Methods Enzymol. 11*, 361.
Shinoda, T., Tatani, K., and Putnam, F. N. (1970), *J. Biol. Chem. 245*, 4463.
Spande, T. F., Witkop, B., Degani, Y., and Patchornik, A. (1970), *Advan. Protein Chem. 24*, 97.
Spiro, G. R. (1966), *Methods Enzymol. 8*, 3.
Steers, E., Jr., Croven, G. R., Anfinsen, C. B., and Bethune, J. L. (1965), *J. Biol. Chem. 240*, 2478.
Wagh, P. V., Bornstein, I., and Winzler, R. J. (1969), *J. Biol. Chem. 244*, 658.
Yamashina, I. (1965), *Acta Chem. Scand. 10*, 1666.
Yemm, C. W., and Cocking, E. C. (1967), *Methods Enzymol. 11*, 325.
Yphantis, D. (1960), *Ann. N. Y. Acad. Sci. 88*, 586.
Zuber, H. (1968), *Hoppe Seyler's Z. Physiol. Chem. 349*, 1337.

Polyadenylic Acid Sequences on 3′ Termini of Vaccinia Messenger Ribonucleic Acid and Mammalian Nuclear and Messenger Ribonucleic Acid†

Robert Sheldon,* Joseph Kates, Dawn E. Kelley, and Robert P. Perry

ABSTRACT: The location of poly(A) on vaccinia mRNA and HeLa n- and mRNA was determined with the use of polynucleotide phosphorylase. This enzyme, which phosphorylizes RNA in a 3′→5′ direction, degraded the poly(A) moiety before the rest of the RNA. When the poly(A) was first removed from the RNA by ribonuclease digestion, then submitted to phosphorolysis, virtually all the poly(A) was degraded. This indicated that all poly(A) sequences contained 3′-hydroxyls, which are characteristic of RNA 3′ termini. The same results were obtained when the poly(A) was degraded by a mammalian 3′-exoribonuclease which hydrolyzes only RNA chains bearing 3′-hydroxyls. The localization of poly(A) on the 3′ terminus of vaccinia mRNA was also supported by the kinetics of addition of the poly(A) to viral RNA being synthesized *in vitro*.

The existence of poly(A) sequences in eukaryotic cells has been recognized for several years (Hadjivassiliou and Brawerman, 1966; Edmonds and Caramela, 1969). The covalent attachment of poly(A) to messenger RNA (mRNA) was demonstrated recently, first for vaccinia mRNA (Kates and Beeson, 1970) and subsequently for other viral (Philipson *et al.*, 1971; Lai and Duesberg, 1972; Weinberg, *et al.*, 1972)[1] and animal cell nuclear and mRNAs (Kates, 1970; Lim and Canellakis, 1970; Darnell *et al.*, 1971b; Edmonds *et al.*, 1971; Lee *et al.*, 1971). Presently very little is known about how the poly(A) is synthesized, or what function it serves. To shed light on these problems it is important to know where on the RNA the poly(A) sequence is located.

In 1970, Kates reported an experiment which indicated that poly(A) is on the 3′ end of vaccinia mRNA. Poly(A), when released from the RNA by T_1 and pancreatic ribonuclease digestion, contained 1 adenosine and 0.2 adenosine tetraphosphate per 110 adenosine 3′-monophosphates. Because the presence of the adenosine could be accounted for only if the poly(A) was 3′ terminal, it appeared that most poly(A) sequences were on the 3′ ends of RNA. The low amounts of adenosine tetraphosphate might have been due to some poly(A) on the 5′ end of RNA, or due to the existence in that system of some free poly(A). Further experiments were needed to clear up this ambiguity.

While the work presented below was in progress, the following information regarding poly(A) location was reported.

† From the Department of Molecular, Cellular, and Developmental Biology (R. S.) and the Department of Chemistry (J. K.), University of Colorado, Boulder, Colorado 80302, and from the Institute for Cancer Research (D. E. K. and R. P. P.), Philadelphia, Pennsylvania 19111. *Received April 24, 1972.* Supported by PHS Grant 1 R01 AI 08413 to J. K. and by NSF Grant GB 15397 to R. P. P.

[1] Submitted for publication.