The Amino Acid Sequence of Peptides from the Cross-Linking Region of Rat Skin Collagen*

Andrew H. Kang,† Paul Bornstein, and Karl A. Piez

ABSTRACT: The amino acid sequences of two cyanogen bromide derived peptides from the cross-linking region at the NH_2-terminal ends of the $\alpha 1$ and $\alpha 2$ chains of rat skin collagen have been established. The sequence of the pentadecapeptide from $\alpha 1$, $\alpha 1$-CB1, was found to be Gly-Tyr-Asp-Glu-Lys-Ser-Ala-Gly-Val-Ser-Val-Pro-Gly-Pro-Hse. The corresponding tetradecapeptide from $\alpha 2$, $\alpha 2$-CB1, was found to have the sequence PCA-Tyr-Ser-Asp-Lys-Gly-Val-Ser-Ala-Gly-Pro-Gly-Pro-Hse. These sequences are not typical of collagen and presumably cannot assume the helical configuration characteristic of the rest of the molecule. These findings may account for the selected susceptibility of the NH_2-terminal region of native collagen to cleavage by a variety of proteolytic enzymes and by cyanogen bromide, and are consistent with the special role of this region as a site of cross-link formation through the lysyl side chains.

Recent studies have led us to conclude that an intramolecular interchain cross-link in rat skin collagen is derived from the side chains of two lysyl residues on adjacent polypeptide chains in the triple-stranded molecule (Bornstein *et al.*, 1966a; Bornstein and Piez, 1966). The regions of the chains involved in the formation of these cross-links were isolated as peptides by cleavage of single-chain ($\alpha 1$ and $\alpha 2$) and double-chain (β_{12}) components at methionyl residues with cyanogen bromide (CNBr)[1] followed by chromatography on phosphocellulose. A pentadecapeptide was obtained from the $\alpha 1$ chain and a similar tetradecapeptide was obtained from the $\alpha 2$ chain. Each existed in two forms, a lysine-containing form ($\alpha 1$-CB1 and $\alpha 2$-CB1) and a lysyl-derived aldehyde-containing form ($\alpha 1$-CB1a and $\alpha 2$-CB1a).[2] The aldehydes were identified as the δ-semialdehyde of α-aminoadipic acid in peptide linkage. The cross-link itself may result from a condensation of two aldehydes. Isotope incorporation studies were consistent with the suggestion that specific lysyl residues serve as precursors of the cross-link through the formation of lysine-derived aldehydes (Piez *et al.*, 1966). In the soluble collagen preparations that were examined only one cross-link per β component was identified, and on the basis of limited cleavage of native collagen with chymotrypsin, trypsin, and CNBr, this bond, and therefore the amino acid sequence surrounding it, was placed at the NH_2-terminal end of the molecule (Bornstein *et al.*, 1966b). The unusual amino acid composition of the NH_2-terminal region of collagen, compared to the whole molecule, and its role in the formation of covalent cross-links prompted us to investigate its amino acid sequence.

Materials and Methods

Preparation of Collagen and Its Fractionation into α Chains and β Components. Acetic acid and neutral salt-extracted collagens were prepared from the skins of 100–150-g Sprague–Dawley rats as described previously (Kang *et al.*, 1966a). α chains and β components were obtained by carboxymethylcellulose chromatography of the heat-denatured collagen preparations (Piez *et al.*, 1963; Bornstein and Piez, 1966). These techniques yield $\alpha 1$ chains in a high degree of purity. $\alpha 2$ chains can also be obtained free of noncollagenous impurities but usually contain approximately 5% β_{12}.

Preparation of Peptides $\alpha 1$-CB1, $\alpha 1$-CB1a, $\alpha 2$-CB1, and $\alpha 2$-CB1a. $\alpha 1$ and $\alpha 2$ chains were cleaved with CNBr in 0.1 N HCl at 30° for 4 hr. After removal of the solvent and reagent by lyophilization the digests were chromatographed on phosphocellulose and the individual peptides were desalted on Bio-Gel P-2. The preparation of these peptides and the criteria for their purity have been described in detail in a preceding communication (Bornstein and Piez, 1966). The procedures reported here were performed on preparations which were at least 90% pure as judged by amino acid analyses. In most instances amino acid containing contaminants accounted for less than 5% of the

* From the Laboratory of Biochemistry, National Institute of Dental Research, National Institutes of Health, Bethesda, Maryland. *Received December 22, 1966.* A portion of this work was presented at the 1966 meeting of the Federation of American Societies for Experimental Biology, Atlantic City, New Jersey (Kang *et al.*, 1966b).

† Present address: Peter Bent Brigham Hospital, Boston, Mass.

[1] Abbreviations used: CNBr, cyanogen bromide; PCA, pyrrolidone-5-carboxylic acid; Hse, homoserine; LAP, leucine aminopeptidase; FDNB, fluorodinitrobenzene; PTH, phenylthiohydantoin.

[2] The symbol CB is used to designate a peptide derived from a protein chain by cyanogen bromide cleavage. Hence, $\alpha 1$-CB1 and $\alpha 2$-CB1 are peptides obtained from the $\alpha 1$ and $\alpha 2$ chains, respectively.

preparation.

Column Chromatography. Phosphocellulose chromatography of tryptic digests of α1-CB1 was carried out at 40° in 0.001 M sodium acetate, pH 3.8, using a linear gradient of NaCl from 0 to 0.3 M over a volume of 820 ml. The procedure was identical with that described for CNBr digests of α1 and α2 (Bornstein and Piez, 1966). Molecular sieve chromatography was performed using either Bio-Gel P-2, 200–400 mesh (Bio-Rad Laboratories), or Sephadex G-25, fine beads (Pharmacia). The Bio-Gel P-2 was packed in a 1.8 × 60 cm column and equilibrated with ammonium propionate buffer, pH 4.5 (Bornstein and Piez, 1966). Fractions of 2.5 ml were collected at a flow rate of 15 ml/hr. The dimensions of the Sephadex column were 1.8 × 90 cm and 2-ml fractions were collected employing the same buffer and flow rate. The optical density of the effluent was monitored at 230 mμ. Cation-exchange chromatography was performed on a 0.4 × 20 cm column of Dowex 50-X2 (Bio-Rad, 200–400 mesh). The resin was washed successively with 2 N sodium hydroxide, water, 2 N HCl, water, and finally with 0.2 M (in pyridine) pyridine acetate buffer, pH 3.1. The peptide samples were applied to the column in 0.5 ml of the pyridine acetate buffer. After application of the sample, 40 ml of pH 3.1 buffer was passed through the column followed by a linear gradient composed of equal volumes (240 ml) of the starting buffer and 2.23 M (in pyridine) pyridine acetate buffer, pH 5.4. Fractions of 1.5 ml were collected at a flow rate of 7 ml/hr. The fractions (0.2 ml of each one) were analyzed with ninhydrin. Fractions comprising the peak tubes in the elution pattern were pooled and lyophilized.

Paper Chromatography. In most cases peptides fractionated by column chromatography were further purified by descending chromatography on Whatman No. 3MM paper. Solvent systems used were: 1-butanol–acetic acid–water (4:1:1), 1-butanol saturated with water, and 1-butanol–pyridine–water (1:1:1). The peptides were located by means of the cadmium ninhydrin reagent (Blackburn, 1965). With this reagent NH_2-terminal glycyl or seryl peptides tend to give a yellow color whereas other peptides stain orange or pink. Tyrosine-containing peptides were located with the Pauly reagent. When paper chromatography indicated purity of peptides isolated by column chromatography, the peptides eluted from the columns were used directly in subsequent analyses. Otherwise, peptides eluted from paper with 1% acetic acid were used.

Amino Acid Analysis. Analyses were performed on a single column automatic analyzer modified for high-speed analysis (Miller and Piez, 1966). Acid hydrolysis was performed in constant-boiling HCl under nitrogen at 108° for 24 hr. Previously determined corrections for hydrolytic loss of serine and tyrosine and incomplete release of valine (Piez *et al.*, 1960) were employed. The analyses of leucine aminopeptidase (LAP) and carboxypeptidase A digests were performed by applying the lyophilized digests, dissolved in 1 ml of citrate buffer, pH 2.9, directly to the analyzer column.

NH_2-Terminal Analyses. NH_2-terminal analyses were performed on the CNBr-produced peptides both by the fluorodinitrobenzene (FDNB) (Fraenkel-Conrat *et al.*, 1955) and the phenyl isothiocyanate methods (Fraenkel-Conrat *et al.*, 1955; Sjöquist, 1960). The procedures used have been described (Piez *et al.*, 1966). In most instances in which Edman degradation was used the results were confirmed by an amino acid analysis of the residual peptide.

Acetyl Group Determinations. Acetyl group determinations were performed by the gas chromatographic procedure of Pisano and Peyton (J. J. Pisano and M. Peyton, in preparation). The peptide sample, usually 0.1–0.4 μmole, was hydrolyzed with 0.25 ml of 6 N H_2SO_4 for 16 hr at 108° in a tube sealed under nitrogen. After hydrolysis 100 mg of anhydrous sodium sulfate was added and the liberated acetate was isolated by five extractions with two volumes of diethyl ether taking care not to contaminate the ether with the aqueous layer (Ward *et al.*, 1966). The combined ether extracts were adjusted to an apparent pH of 10 with 0.01 N NaOH in 95% methanol and then evaporated to dryness. H_2O (100 μl) and 10 mg of washed Dowex 50-X8, 200–400 mesh, in the hydrogen form, were added to the dried samples. The mixture was shaken for 1 min to liberate acetic acid from the sodium salt. A 1–5-μl aliquot of the aqueous solution was injected into the gas chromatograph. A Barber-Colman Model 15 instrument equipped with a hydrogen flame ionization detector and a glass 6-ft U column was used. The solid support was a polyaromatic polymer in beaded form, 100–120 mesh, prepared from ethylvinylbenzene with divinylbenzene as a cross-linker (Hollis, 1966). It was purchased as Porapak Q from Waters Associates, Inc., Framingham, Mass. The beads were washed on a scintered-glass funnel with ethanol and coated by the filtration technique (Horning *et al.*, 1959) with a chloroform solution containing 10 g of Carbowax 20M–terephthalic acid/100 ml (Applied Science Laboratories, Inc., State College, Pa.). The column was conditioned at 230° for 1 hr. Operating conditions were: carrier gas argon, 16 psi; flash heater 200°; column 192°; and detector cell 240°. Solutions containing 0.1–1.0 μg/μl of acetic acid were used to obtain a standard curve.

Hexose Determinations. Determinations for glucose and galactose were performed using the Glucostat and Galactostat reagents (Worthington). These reagents combine the actions of glucose and galactose oxidases together with that of a peroxidase. In the presence of an appropriate chromagen the absorbance (at 400 mμ for glucose and 425 mμ for galactose) is proportional to hexose concentration. These methods have the advantage of specificity and sensitivity permitting the detection of as little as 0.05 μmole of glucose or galactose. A solution containing 50 μg of glucose and 50 μg of galactose/ml was used to obtain a standard curve. In each instance, good reproducibility and a straight-line relation between concentration of hexose and optical density were obtained over the range

used (10–50 μg of hexose/ml of sample). Determinations were performed both before and after mild acid hydrolysis (1 N HCl, 108° for 60 min) of peptides in order to detect hexoses present in glycosidic linkage.

Digestion with Trypsin and Chymotrypsin. CNBr-produced peptides were cleaved by trypsin (two times crystallized, Worthington) or α-chymotrypsin (three times crystallized, Worthington) in 0.2 M NH_4HCO_3, pH 7.8, at 37° for 4 hr. An enzyme:substrate molar ratio of 1:50 was employed, and in the case of the tryptic digests, $CaCl_2$ was added to make a 0.005 M solution. The reaction was terminated by lowering the pH to 4 with 2 N acetic acid. The concentrations of peptides in these digestions as well as in digestions with pepsin, pronase, and carboxypeptidase A ranged from 0.1 to 1 μmole/ml of reaction mixture.

Digestion with Pepsin. The digestion with pepsin (two times crystallized, Worthington) was carried out in 0.01 M HCl at 37° for 24 hr at an enzyme:substrate molar ratio of 1:50. The reaction was terminated by lyophilization.

Digestion with Pronase. The digestion with pronase (B grade, Calbiochem) was carried out in 0.2 N NH_4HCO_3, pH 8.0, at 37° for 48 hr at an enzyme:substrate molar ratio of 1:50. Toluene (1 drop) was added to prevent bacterial growth.

Digestion with Leucine Aminopeptidase. The digestion was performed in 0.2 M NH_4HCO_3 buffer, pH 8.0, containing 0.005 M $MgSO_4$. LAP (Worthington) was first dialyzed against a large volume of the buffer for 4 hr in the cold, and was activated for 1 hr at 37° prior to digestion. The enzyme:substrate molar ratio ranged from 1:7000 to 1:50, as specified in the text in each case.

Digestion with Carboxypeptidase A. Crystalline carboxypeptidase A (DFP treated, Worthington) was dissolved in 10% NH_4HCO_3, pH 8.0, and quantitated by determining the optical density at 278 mμ using $E_{1\,cm}^{0.1\%}$ 1.94. Digestions were performed in 0.2 M NH_4HCO_3, pH 8.0, at 37°. The enzyme:substrate molar ratios and the duration of the digestions are specified below for individual experiments.

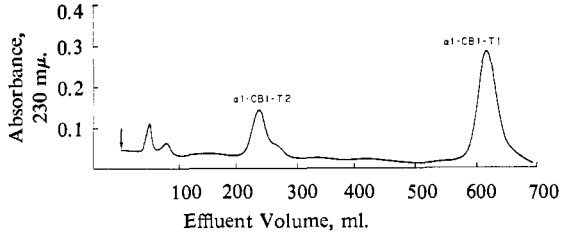

FIGURE 1: Phosphocellulose elution pattern of a tryptic digest of α1-CB1. Chromatography was performed at pH 3.8, 40°. The arrow indicates placement of the sample (1 μmole) dissolved in 10 ml of starting buffer. Two peptides, α1-CB1-T1 and α1-CB1-T2, are identified. The enzyme is not eluted from the column under the conditions of chromatography.

Degradation of Peptides by the Edman Procedure. Peptides were degraded sequentially from the amino terminus by the procedure of Edman and Sjöquist (1956). PTH amino acids were identified by paper chromatography and quantitated after elution by measuring the absorbance at appropriate wavelengths (Sjöquist, 1960). In some instances confirmation was obtained by an amino acid analysis of the residue.

Results

NH_2-Terminal Analysis of Peptides α1-CB1, α1-CB1a, α2-CB1, and α2-CB1a. The presence of free α-amino groups on peptides α1-CB1, α1-CB1a, α2-CB1, and α2-CB1a was investigated by dinitrophenylation and Edman degradation. In seven analyses an average of 1.0 equiv of PTH-glycine (range 0.85–1.1) was obtained/mole of α1-CB1. DNP-glycine was apparently destroyed almost completely during the hydrolysis and only trace amounts were detected. No NH_2-terminal residue except in trace amounts could be detected in α1-CB1a, α2-CB1, or α2-CB1a by either method. However, N^{ϵ}-DNP-lysine was identified in the water-soluble fraction after dinitrophenylation of α1-CB1 and α2-CB1. These results were confirmed by amino acid analyses of the residues after Edman degradation.

Analysis for Amino-Blocking Groups. In view of the apparent absence of available NH_2-terminal groups in α1-CB1a, α2-CB1, and α2-CB1a, and of published reports indicating the presence of acetyl groups in collagen (Hörmann *et al.*, 1965; Hörmann and Joseph, 1965), the possibility that acetyl groups might be blocking the NH_2-terminal groups was considered. Quantities of these peptides ranging from 0.1 to 0.4 μmole were analyzed for acetyl groups. None was found in any of the peptides. As controls, similar amounts of *N*-acetylamino acids, ethyl acetate, sodium acetate, and acetic acid were analyzed. In each instance, the expected amount of acetic acid was recovered.

The possibility that NH_2-glycosidic groups might be responsible for the absence of free NH_2 terminals in these peptides, as suggested for insoluble collagen by Chandrajan and Bose (1965), was also considered. Since virtually all the carbohydrate in soluble vertebrate collagens can be accounted for by glucose and galactose (see Seifter and Gallop, 1966), the enzymes glucose oxidase and galactose oxidase were used. α1-CB1a (0.25 μmole) (determined by amino acid analysis on an aliquot) was subjected to mild acid hydrolysis (see Methods). After removal of HCl by flash evaporation, the partial acid hydrolysate as well as an equal quantity of the unhydrolyzed peptide were tested with glucose and galactose oxidases. As a control 0.6 μmole of α1-CB1 was similarly divided and treated. Neither glucose nor galactose was found in any sample. If equimolar quantities of hexose had been present, 0.15 μmole or 27 μg of hexose in α1-CB1 or 0.125 μmole or 23 μg of hexose in α1-CB1a should have been detected. Hexose determinations were not performed on peptides α2-CB1 and α2-CB1a

TABLE I: Amino Acid Composition and R_F Values of Tryptic and Peptic Peptides of α1-CB1.[a]

		α1-CB1-					
	α1-CB1	T1	T1-P1	T1-P2	T2	T2-P1	T2-P2
Aspartic acid	1.0	1.0	–	0.9	–	–	–
Serine	2.0	–	–	–	2.0	0.9	1.0
Homoserine	0.9	–	–	–	0.9	–	1.0
Glutamic acid	1.0	1.0	–	1.0	–	–	–
Proline	1.8	–	–	–	1.8	–	1.8
Glycine	3.2	1.1	1.0	–	2.1	1.0	1.1
Alanine	1.1	–	–	–	1.0	1.0	–
Valine	1.8	–	–	–	2.1	–	2.0
Tyrosine	1.2	0.9	1.0	–	–	–	–
Lysine	1.0	1.0	–	1.0	–	–	–
R_F values		0.09	0.48	0.08	0.52	0.27	0.49
Color with ninhydrin		Orange	Yellow	Pink	Yellow	Yellow	Pink

[a] Composition expressed as residues per peptide. Homoserine includes homoserine lactone. A dash indicates 0.1 residue or less. R_F values were obtained using descending paper chromatography with 1-butanol–acetic acid–water (4:1:1). A portion of these data has been published (Bornstein and Piez, 1966).

since amino acid sequence studies (see below) provided an adequate explanation for the absence of free NH_2 terminals in these peptides.

Amino Acid Sequence of α1-CB1. Because of the possible complications arising from the presence of the lysyl-derived aldehydes in α1-CB1a and α2-CB1a, all enzymatic degradations and subsequent sequence studies were performed on the lysine-containing precursor forms of these peptides, α1-CB1 and α2-CB1. The chromatography of the tryptic peptides of α1-CB1 on phosphocellulose is shown in Figure 1. Two peptides, T1 and T2, were obtained. α1-CB1-T1 was found to be a pentapeptide containing a lysyl residue and α1-CB1-T2 a decapeptide containing a residue of homoserine (lactone). The amino acid composition of these peptides and their R_F values in 1-butanol–acetic acid–water are given in Table I.

Peptide α1-CB1-T1 (Residues 1–5). GLY-TYR-ASP-GLU-LYS. Since α1-CB1-T1 results from the tryptic cleavage of α1-CB1, lysine must be in the COOH-terminal position and this peptide must constitute the NH_2-terminal sequence of α1-CB1. This is in agreement with the finding of homoserine in α1-CB1-T2 since α1-CB1, which results from CNBr cleavage of a methionyl bond, terminates in homoserine. Hydrolysis of α1-CB1-T1 with LAP for 24 hr at an enzyme:substrate molar ratio of 1:1000 gave Gly 1.02 and Tyr 0.28, whereas hydrolysis at a ratio of 1:70 for 24 hr gave Gly 1.02, Tyr 1.08, Asp 0.37, Glu 0.37 and Lys 0.37, indicating the sequence to be Gly-Tyr-(Asp, Glu)-Lys. These results as well as the results of subsequent LAP digestions are expressed as moles of amino acids per mole of the substrate peptide. No asparagine or glutamine was found in LAP digests.

A peptic digest of α1-CB1-T1 was fractionated on Dowex 50-X2. Three peptides were eluted of which one was found to be the uncleaved peptide. The amino acid compositions and R_F values of the two peptic fragments of α1-CB1-T1 are given in Table I. Edman degradation of α1-CB1-T1-P2 showed the sequence to be Asp-Glu-Lys. No PTH derivatives of asparagine or glutamine were found. Further, ammonia was found in far less than stoichiometric amounts in acid hydrolysates of α1-CB1-T1-P2. The position of the lysyl residue as number 5 in α1-CB1 was confirmed by qualitative Edman degradation of α1-CB1 through five residues. The results of this sequential degradation substantiated the findings obtained with LAP and pepsin.

It is of interest that chymotrypsin cleaved the tyrosyl bond in α1-CB1 very slowly if at all. A paper chromatogram of a chymotryptic digest of α1-CB1-T1 (enzyme:substrate molar ratio 1:25, 37° for 24 hr) revealed only faint ninhydrin-positive spots corresponding to Gly-Tyr and Asp-Glu-Lys. Similarly, virtually no cleavage was detected after Bio-Gel P-2 chromatography of a chymotryptic digest of α1-CB1. No evidence was obtained for the possibility that the tyrosyl residue might exist in an altered form such as tyrosyl *O*-sulfate as occurs in fibrinogen (Bettelheim, 1954). The ultraviolet absorption spectrum of α1-CB1 was characteristic of a normal tyrosine with a peak at 275 mμ in 0.1 N HCl and a shift to a higher wavelength in 0.1 N NaOH.

Peptide α1-CB1-T2 (Residues 6–15). SER-ALA-GLY-VAL-SER-VAL-PRO-GLY-PRO-HSE. NH_2-terminal analyses by Edman degradation and the FDNB method showed serine to be the NH_2-terminal residue. A peptic digest of α1-CB1-T2 could not be resolved by

TABLE II: Sequential Edman Degradation of Peptide α1-CB1-T2-P2.

Degradation	PTH-Amino Acid	Residue Analysis				
		Serine	Valine	Proline	Glycine	Homoserine
Original Peptide		1	2	2	1	1
Step 1	Valine					
Step 2	Serine	0.06	1.07	1.86	1.03	1.10
Step 3	Valine					
Step 4	Proline	0.04	0.15	1.20	1.03	1.10
Step 5	Glycine					
Step 6	Proline					

Dowex 50-X2 chromatography. However, two well-separated ninhydrin-positive spots were observed after paper chromatography with 1-butanol–acetic acid–water. The amino acid composition and R_F values of these peptic fragments are given in Table I.

Edman degradation of α1-CB1-T2-P1 indicated the amino acid sequence to be Ser-Ala-Gly. Hydrolysis with LAP at a molar ratio of 1:1000 for 3 hr at 37° gave Ser 0.78, Ala 0.40, and Gly 0.40. The amino acid sequence of peptide α1-CB1-T2-P2 was established by means of Edman degradation. The results are summarized in Table II.

Confirmation of part of the sequence of α1-CB1-T2-P2 was obtained by experiments with LAP. Hydrolysis with LAP at a ratio of 1:1000 for 30 min at 37° yielded Val 0.80 and Ser 0.40. No amino acids other than valine and serine could be liberated by LAP. Thus, hydrolysis at a 1:100 ratio for 3 hr yielded Val 1.00, Ser 0.98; after 24 hr, the values were identical. No amino acids were liberated by the action of carboxypeptidase A. The limited activity of LAP and the lack of activity of carboxypeptidase A may be accounted for by the presence of penultimate prolyl residues. α1-CB1-T2-P2 was not cleaved by bacterial collagenase.

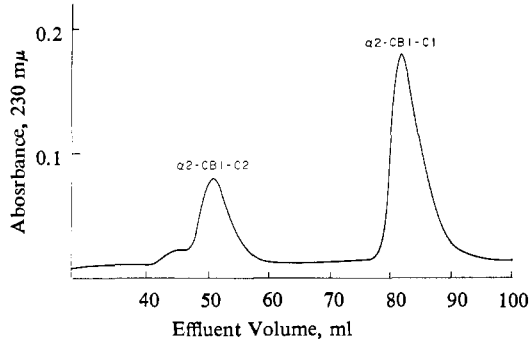

FIGURE 2: Sephadex G-25 elution pattern of chymotryptic digest of α2-CB1. The sample (1 μmole) was dissolved in 2 ml of ammonium propionate buffer, pH 4.5. Two chymotryptic peptides, α2-CB1-C1 and α2-CB1-C2, are identified.

Amino Acid Sequence of α2-CB1. Figure 2 depicts the elution diagram of a chymotryptic digest of α2-CB1 on Sephadex G-25. Two peptides, C1 and C2, were isolated. Peptide α2-CB1-C2 contained a lysyl residue and was further digested with trypsin and the products were separated on Bio-Gel P-2. Two peptides, α2-CB1-C2-T1 and α2-CB1-C2-T2, were separated from each other and from the enzyme by this procedure. The amino acid composition and the R_F values of these peptides are given in Table III.

Peptide α2-CB1-C1 (Residues 1–2). PCA-TYR. Despite its considerably smaller size relative to α2-CB1-C2, the extinction of this peptide at 230 mμ was greater (Figure 2) presumably owing to the presence of tyrosine.

TABLE III: Amino Acid Composition and R_F Values of Chymotryptic and Tryptic Peptides of α2-CB1.[a]

	α2-CB1	α2-CB1-			
		C1	C2	C2-T1	C2-T2
Aspartic acid	1.0	–	1.1	1.0	–
Serine	2.0	–	2.0	1.0	1.0
Homoserine	1.0	–	0.9	–	0.9
Glutamic acid	1.0	1.0	–	–	–
Proline	1.9	–	1.9	–	1.9
Glycine	3.1	–	3.0	–	3.0
Alanine	1.0	–	1.0	–	1.0
Valine	1.0	–	1.0	–	1.1
Tyrosine	1.1	1.1	–	–	–
Lysine	0.9	–	1.0	1.0	–
R_F values		0.73			0.49
Color with ninhydrin		None			Pink

[a] Composition expressed as residues per peptide. Homoserine includes homoserine lactone. A dash indicates 0.1 residue or less. R_F values were obtained using descending chromatography with 1-butanol–acetic acid–water (4:1:1).

Its retardation on Sephadex can also be explained on this basis. Peptide α2-CB1-C1 did not react with the ninhydrin reagent and no free NH_2 terminals were detected by the phenyl isothiocyanate or FDNB methods. It was located on paper by the Pauly reagent. This finding plus the presence of homoserine in α2-CB1-C2 established α2-CB1-C1 as the NH_2-terminal portion of α2-CB1. Hydrolysis of α2-CB1-C1 with carboxypeptidase A at an enzyme:substrate molar ratio of 1:50 for 24 hr yielded Tyr 1.08, and no glutamic acid or glutamine. Consequently, the possibility of ar NH_2-terminal pyrrolidone-5-carboxylic acid (PCA) residue was examined. The paper chromatographic properties of α2-CB1-C1 were compared with those of a sample of authentic PCA-tyrosine in three solvent systems, 1-butanol–acetic acid–water (4:1:1), 1-butanol saturated with water, and 1-butanol–pyridine–water (1:1:1). Approximately 0.1 μmole of each peptide was applied to Whatman No. 3MM paper and the sheets were developed for 12 hr using descending chromatography. Peptides were located by spraying with the Pauly reagent. Comparisons of the R_F values of peptide α2-CB1-C1 with PCA-tyrosine showed them to be identical in all three systems. Since cyclization may well have occurred under the acidic conditions used in the preparation of α2-CB1, the presence of NH_2-terminal PCA *in vivo* cannot be assumed.

Peptide α2-CB1-C2-T1 (Residues 3–5). SER-ASP-LYS. Edman degradation established serine as the NH-terminal residue. The second Edman degradation gave PTH-Asp. Hydrolysis with LAP at an enzyme:substrate molar ratio of 1:100 for 24 hr at 37° gave Ser 1.00, Asp 0.90, and Lys 0.90. No asparagine was observed. The presence of far less than a stoichiometric quantity of ammonia after acid hydrolysis of α2-CB1-C2-T1 further excluded the presence of asparagine in this peptide.

Peptide α2-CB1-C2-T2 (Residues 6–14). GLY-VAL-SER-ALA-GLY-PRO-GLY-PRO-HSE. Edman degradation of this peptide established glycine as the NH_2-terminal residue, followed by valine. Hydrolysis with LAP at an enzyme:substrate molar ratio of 1:1000 for 1 hr at 37° gave Gly 0.36, Val 0.28, Ser 0.22, Ala 0.08; after 5 hr: Gly 0.84, Val 0.79, Ser 0.74, and Ala 0.58. At a ratio of 1:250, 24-hr incubation at 37°, LAP hydrolysis gave Gly 0.95, Val 0.95, Ser 0.97, and Ala 0.97. Thus the partial sequence could be deduced at this point to be Gly-Val-Ser-Ala(Gly_2,Pro_2)-Hse. Since α2-CB1 resulted from the CNBr cleavage of a methionyl bond, the peptide must terminate in homoserine. In order to investigate the sequence of α2-CB1-C2-T2 further, the peptide was digested with pronase and the resulting peptides were fractionated on Dowex 50-X2. Two peaks were obtained. The material in each was chromatographed on Whatman No. 3MM paper using 1-butanol–acetic acid–water (4:1:1). Four ninhydrin-positive spots were obtained from one peak and two ninhydrin-positive spots from the other peak. The corresponding peptides were isolated by preparative paper chromatography and characterized (Table IV).

Edman degradation of two peptides, Pro2 and Pro-5, was sufficient to establish the sequence of α2-CB1-C2-T2. These results are summarized in Tables V and VI. Edman degradations on the remaining pronase-derived peptides (data not presented) confirmed this sequence (see Figure 3).

Discussion

The data presented here permit the establishment of the complete sequences of peptides α1-CB1 and α2-CB1. These sequences together with the positions of cleavage by the proteolytic enzymes employed are summarized in Figure 3. The evidence that these peptides represent the NH_2-terminal sequences of the α1 and α2 chains, respectively, has been presented in a preceding communication (Bornstein *et al.*, 1966b). The finding of NH_2-terminal glycine in α1-CB1 and the absence of free NH_2-terminals in α1-CB1a, α2-CB1, and α2-CB1a together with identical findings in the α1 and α2 chains from which these peptides were

TABLE IV: Amino Acid Composition and R_F Values of Peptides Derived from α2-CB1-C2-T2 by Pronase Digestion.[a]

	Peak 1, Dowex 50				Peak 2, Dowex 50	
	Pro1	Pro2	Pro3	Pro4	Pro5	Pro6
Serine	–	–	–	–	1.0	–
Homoserine[b]	–	1.0	–	–	–	–
Proline	2.1	2.1	1.1	–	–	–
Glycine	2.0	2.1	1.0	1.0	2.0	1.0
Alanine	0.9	1.0	0.9	–	1.0	–
Valine	–	–	–	1.0	1.0	–
R_F values	0.25	0.33	0.39	0.58	0.36	
Color with ninhydrin	Pink	Pink	Pink	Yellow	Yellow	Pink

[a] Composition expressed as residues per peptide. R_F values were obtained using descending paper chromatography with 1-butanol–acetic acid–water (4:1:1). [b] Homoserine includes homoserine lactone.

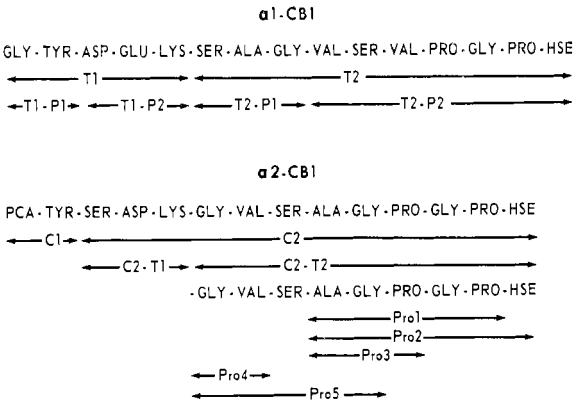

FIGURE 3: Amino acid sequences of α1-CB1 and α2-CB1 and positions of cleavage by the proteolytic enzymes: T, trypsin; C, chymotrypsin; P, pepsin; and Pro, pronase. PCA constitutes the NH_2-terminal residue of α2-CB1. Hse results from cyanogen bromide cleavage at preexisting methionyl residues.

derived (Piez *et al.*, 1966) serve to confirm the NH_2-terminal position of these peptides in the parent chains.

The similarity between the amino acid compositions and sequences of α1-CB1 and α2-CB1 is of interest. The sequences of the two peptides can be arranged so that 9 of the 14 or 15 amino acids are identical, namely tyrosine, lysine, the sequence Gly-Val-Ser, and the terminal sequence Pro-Gly-Pro-Hse. These findings provide support for the concept that the α1 and α2 chains of rat skin collagen arose from a common precursor.

The nature of the amino blocking group present in α1-CB1a and in the parent α1 chain remains unknown. We have been unable to identify an acetyl group, glucose, or galactose in the peptide. The association of the disappearance of NH_2-terminal glycine with the appearance of a lysyl-derived aldehyde in position 5 suggests the possibility that an internal Schiff's base may be present in α1-CB1a. In lathyritic collagen, in which the conversion of the lysyl residues in position 5 of both α1 and α2 to the lysyl-derived δ-semialdehyde of α-aminoadipic acid is inhibited, NH_2-terminal glycine persists in the α1 chain (Bornstein and Piez, 1966; Piez *et al.*, 1966).

TABLE V: Sequential Edman Degradation of Peptide α2-CB1-C2-T2-Pro2.

Degradation	PTH-Amino Acid	Residue Analysis: Alanine	Glycine	Proline	Homoserine
Original peptide		1	2	2	1
Step 1	Alanine				
Step 2	Glycine				
Step 3	Proline	0.05	1.20	1.20	1.10
Step 4	Glycine				

TABLE VI: Sequential Edman Degradation of Peptide α2-CB1-C2-T2-Pro5.

Edman Degradation	PTH-Amino Acid	Residue Analysis: Glycine	Valine	Serine	Alanine
Original peptide		2	1	1	1
Step 1	Glycine	0.96	1.05	1.01	1.06
Step 2	Valine	1.05	0.19	1.06	1.00
Step 3	Serine				
Step 4	Alanine				

The amino acid composition and sequence of the NH_2-terminal region of rat skin collagen are unusual in several respects. This region is more acidic than collagen, lacks hydroxyproline, and does not contain glycine in every third position as does the main body of the collagen molecule. By present concepts it therefore cannot assume the helical configuration characteristic of collagen. Its unusual structure probably accounts for its susceptibility in the undenatured protein to proteolytic enzymes and to CNBr (Bornstein *et al.*, 1966b). The selective enzymatic oxidation of the lysyl residues near the NH_2 terminus of collagen may also be determined in part by the configuration of this region.

It is likely that the NH_2-terminal region corresponds in part to that portion of the molecule released from native collagen by pepsin (Rubin *et al.*, 1963, 1965) and other proteolytic enzymes (Drake *et al.*, 1966). It is also likely that the ninhydrin-negative aldehyde-containing peptide isolated from collagenase digests of carp swim bladder collagen (Rojkind *et al.*, 1964, 1966) originates from a corresponding region of the collagen molecule. It has been shown that the peptides released by the action of proteolytic enzymes on native collagen, termed telopeptides (Schmitt, 1964), are responsible for much of the antigenicity of collagen (Schmitt *et al.*, 1964). The precise structure of the NH_2-terminal region of collagen may, therefore, be of additional importance in the study of the immunogenicity of the protein.

Acknowledgment

The authors are indebted to Dr. John J. Pisano for his assistance in performing the gas chromatographic analyses, Dr. J. E. Folk for his generous gift of PCA-tyrosine, and Mr. Robert Katzen and Mrs. Suzanne Roniger for their able technical assistance.

References

Bettelheim, F. R. (1954), *J. Am. Chem. Soc. 76*, 2838.

Blackburn, S. (1965), *Methods Biochem. Analy. 13*, 1.

Bornstein, P. Kang, A. H., and Piez, K. A. (1966a), *Proc. Natl. Acad. Sci. U. S. 55*, 417.

Bornstein, P., Kang, A. H., and Piez, K. A. (1966b), *Biochemistry 5*, 3803.

Bornstein, P., and Piez, K. A. (1966), *Biochemistry 5*, 3460.

Chandrajan, J., and Bose, S. M. (1965), *Indian J. Biochem. 2*, 127.

Drake, M. P., Davison, P. F., Bump, S., and Schmitt, F. O. (1966), *Biochemistry 5*, 301.

Edman, P., and Sjöquist, J. (1956), *Acta Chem. Scand. 10*, 1507.

Fraenkel-Conrat, H., Harris, J. I., and Levy, A. L. (1955), *Methods Biochem. Analy. 2*, 359.

Hollis, O. L. (1966), *Anal. Chem. 38*, 309.

Horning, E. C., Moscatelli, E. A., and Sweeley, C. C. (1959), *Chem. Ind. (London)*, 751.

Hörmann, H., and Joseph, K. T. (1965), *Biochim. Biophys. Acta 100*, 598.

Hörmann, H., Joseph, K. T., and Von Wilm, M. (1965), *Z. Phys. Chem. 341*, 284.

Kang, A. H., Bornstein, P., and Piez, K. A. (1966b), *Federation Proc. 25*, 716.

Kang, A. H., Nagai, Y., Gross, J., and Piez, K. A. (1966a), *Biochemistry 5*, 509.

Miller, E. J., and Piez, K. A. (1966), *Anal. Biochem. 16*, 320.

Piez, K. A., Eigner, E. A., and Lewis, M. S. (1963), *Biochemistry 2*, 58.

Piez, K. A., Martin, G. R., Kang, A. H., and Bornstein, P. (1966) *Biochemistry 5*, 3813.

Piez, K. A., Weiss, E., and Lewis, M. S. (1960), *J. Biol. Chem. 235*, 1897.

Rojkind, M., Blumenfeld, O. O., and Gallop, P. M. (1964), *Biochem. Biophys. Res. Commun. 17*, 320.

Rojkind, M., Blumenfeld, O. O., and Gallop, P. M. (1966), *J. Biol. Chem. 241*, 1530.

Rubin, A. L., Drake, M. P., Davison, P. F., Pfahl, D., Speakman, P. T., and Schmitt, F. O. (1965), *Biochemistry 4*, 181.

Rubin, A. L., Pfahl, D., Speakman, P. T., Davison, P. F., and Schmitt, F. O. (1963), *Science 139*, 37.

Schmitt, F. O. (1964), *Federation Proc. 23*, 618.

Schmitt, F. O., Levine, L., Drake, M. P., Rubin, A. L., Pfahl, D., and Davison, P. F. (1964), *Proc. Natl. Acad. Sci. U. S. 51*, 493.

Seifter, S., and Gallop, P. M. (1966), *Proteins 4*, 155.

Sjöquist, J. (1960), *Biochim. Biophys. Acta 41*, 20.

Ward, D. N., Coffey, J. A., Ray, D. B., and Lamkin, W. M. (1966), *Anal. Biochem. 14*, 243.