Amino Acid Sequence of Cyanogen Bromide Peptides from the Amino-Terminal Region of Chick Skin Collagen*

Andrew H. Kang† and Jerome Gross

ABSTRACT: The amino acid sequence of three cyanogen bromide peptides derived from the α1 chain of chick skin collagen, α1-CB0, α1-CB1, and α1-CB2, and of one peptide derived from the α2 chain, α2-CB1, has been determined. α1-CB0, a dipeptide, is the NH_2-terminal peptide of α1. α1-CB1 is a septadecapeptide and the next in position. It contains the lysyl residue involved in the formation of the intramolecular cross-link. Next in position is α1-CB2 containing 36 amino acid residues. Thus, the sequence of 55 residues from the NH_2 terminus of the α1 chain of chick skin collagen is now established. α2-CB1 is the NH_2-terminal peptide of the α2 chain and contains 15 amino acid residues. It also contains the precursor lysyl residue for the intramolecular cross-link. The sequence of the three NH_2-terminal peptides of the α1 chain of the chick skin collagen (α1-CB0, α1-CB1, and α1-CB2) is quite homologous to the known sequence of the corresponding region of rat skin collagen. The only differences are substitution of alanine for serine at the positions 14 and 22 and those related to the degree of hydroxylation of certain prolyl residues. However, the sequence of α2-CB1 of chick skin collagen is markedly different from that of α2-CB1 of rat skin collagen. The relatively strict preservation of the sequence of the α1 chain between the collagens from two different classes of vertebrates and the comparatively large differences in the sequence of the α2 chain, at least as far as is presently known, suggest the possibility that the α2 chain may subserve a special function in determining the species-specific properties of collagen.

The sequence presented here indicates that the NH_2-terminal regions of the α1 and α2 chains could not exist in the helical forms characteristic of the main body of the native collagen molecule. This is consistent with the known susceptibility of this region of the molecule to a variety of proteolytic enzymes and also with its involvement as a cross-linking site.

Knowledge of the primary structure of collagen is essential in understanding the structure–function relationship of this protein. Furthermore, definitive understanding of this relationship will be required in the investigation of pathogenesis of certain human disorders of connective tissue. Information on the comparative biochemistry of collagens from various classes of vertebrates may, as has been shown in other proteins, significantly contribute toward our understanding of those aspects of primary structure which determine the biologic properties of the protein. In this regard, we have recently reported the separation and characterization of ten CNBr peptides of the α1 chain and six CNBr peptides of the α2 chain of chick skin collagen (Kang *et al.*, 1969a,b) accounting for all of the known amino acids of the two chains. Similar studies on rat skin collagen (Butler *et al.*, 1967; Fietzek and Piez, 1969) and on chick bone collagen (Miller *et al.*, 1969; Lane and Miller, 1969) have also been reported.

The NH_2-terminal regions of the collagen chains are of special interest because of their uniquely different amino acid composition as compared with the main body of the collagen molecule, and their participation in the formation of the intramolecular cross-links (Bornstein and Piez, 1966; Kang *et al.*, 1969c) and probably of at least one of the intermolecular cross-links (A. H. Kang and J. Gross, unpublished data). Furthermore, the immunogenic specificity of collagen may be determined by the structure of the NH_2-terminal region of the molecule (Michaeli, 1969).

In this paper we report the sequence of three α1 CNBr peptides, α1-CB0, α1-CB1 and α1-CB2, which in that order represent the NH_2-terminal sequence of the α1 chain, and the sequence of α2-CB1 which is the NH_2-terminal peptide of the α2 chain of chick skin collagen. The sequence of homologous peptides of rat skin collagen (Kang *et al.*, 1967; Bornstein, 1967, 1969) has been reported.

Experimental Section

Preparation of Collagen and Its Fractionation into α1 and α2. Neutral salt-extracted chick skin collagen was prepared from 3-week-old white Leghorn chick according to the procedure described in a previous communication (Kang *et al.*, 1969c). The α1 and α2 chains were obtained by carboxymethyl-cellulose chromatography of denatured, solubilized collagen (Piez *et al.*, 1963; Kang *et al.*, 1969c).

Preparative of the Peptides, α1-CB0, α1-CB1, α1-CB2, and α2-CB1. The purified α chains were digested with CNBr at 30° under nitrogen for 4 hr in 0.1 N HCl and lyophilized. The dried CNBr digests were fractionated on columns of phosphocellulose. The resulting peptide fractions containing the individual peptides were desalted on Bio-Gel P-2. The details of these procedures and the criteria for the purity

* From the Developmental Biology Laboratory, Departments of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts. *Received September 4, 1969.* This is Publication No. 498 of the Robert W. Lovett Memorial Group for the Study of Diseases Causing Deformities. This investigation was supported by grants from the National Institutes of Health, Bethesda, Md. (AM 3564), and the RGK Foundation.

† Senior Investigator of The Arthritis Foundation at the Massachusetts General Hospital and Harvard Medical School, Boston, Mass.

of the peptides have been reported (Kang *et al.*, 1969c). The work reported here has been performed on preparations which were at least 90% pure.

Enzyme Hydrolyses. Hydrolyses with trypsin (L-(1-tosylamido-2-phenyl)ethyl chloromethyl ketone treated, Worthington) and α-chymotrypsin (three-times recrystallized, Worthington) were performed in 0.2 M NH_4HCO_3 (pH 7.8) at 37° for 4 hr. An enzyme:substrate molar ratio of 1:50 was used. In the case of tryptic digestion, $CaCl_2$ was added to the reaction mixture to the final concentration of 1×10^{-3} M. The reaction was terminated by addition of 2 N acetic acid to pH 4 and lyophilization.

Digestion with pepsin (twice crystallized, Worthington) was carried out in 0.01 N HCl at 37° for 16 hr at an enzyme: substrate molar ratio of 1:50. The reaction was terminated by lyophilization.

Crystalline suspension of papain (Worthington) was dissolved in cold water and quantitated by measuring the optical density at 280 mμ using an $E^{1\%}_{1cm}$ of 25. The digestion was performed in 0.07 M pyridine acetate buffer (pH 5.5) containing 0.07 M mercaptoethanol at 37° for 48 hr. The concentration of enzyme was 2–5% that of the substrate. The digestion was terminated by lyophilization.

The digestion with LAP[1] (Worthington) was performed in 0.2 M NH_4HCO_3 (pH 8.0) containing 5×10^{-3} M $MgSO_4$ at 37°. LAP was first dialyzed against a large volume of the buffer and activated for 1 hr at 37° prior to use. The enzyme: substrate molar ratio ranged from 1:1000 to 1:70 as specified in the text.

Bacterial collagenase (CLSPA, Worthington) was used without further purification. It contained no detectable activity against casein even after incubation at 37° for 24 hr. Digestion was carried out at 37° for 24 hr in 0.2 M NH_4HCO_3 (pH 7.8) containing 1×10^{-3} M $CaCl_2$ at an enzyme concentration of 0.05 mg/ml. The substrate concentration was 4 mg/ml. A drop of toluene was used to retard bacterial growth. The digestion was terminated by lyophilization.

Pyrrolidonyl peptidase[2] was used to remove the NH_2-terminal PCA residues in some instances. The reaction was performed in 0.05 M phosphate buffer (pH 7.3) containing 0.03 M mercaptoethanol and 0.001 M EDTA at 30°. The reaction was terminated by lyophilization.

Digestion with CPA (Worthington) was performed in 0.2 M NH_4HCO_3 buffer (pH 7.8) at 37°. The reaction was terminated by lyophilization and the dried reaction mixture was analyzed for amino acids on an automatic amino acid analyzer.

Column Chromatography. Molecular sieve chromatography on Bio-Gel P-2 or Bio-Gel P-4 (200–400 mesh) was used for separation of some of the peptide fragments. Columns (2 × 60 cm) were packed with the resin slurry, equilibrated with 0.15 M acetic acid. The optical density of the effluent was monitored at 230 mμ.

Cation-exchange chromatography was performed on 0.9 × 50 cm columns of Spherix (Phoenix Instruments) or 0.9 × 25 cm columns of PA-35 (Beckman) using an automatic peptide analyzer (Beckman) equipped with stream-split device. The resins were washed successively with 2 N NaOH, water, 3 N HCl, water, 2 M pyridine, and finally with 0.2 M pyridine acetate (pH 3.0). The columns were jacketed at 50°. The peptide samples were applied to the column in 1 ml of water with the pH adjusted to 2.5 with HCl. Elution was carried out with 70 ml of the starting buffer followed by a linear gradient composed of equal volumes (350 ml) of the starting buffer and 2.0 M pyridine acetate (pH 5.0). A flow rate of 70 ml/hr was used. By the use of a stream-split device, a portion of the effluent was analyzed continuously for ninhydrin reactivity and the remainder was collected in 2-min fractions. Appropriate fractions were pooled and lyophilized.

Phosphocellulose chromatography of some of the enzymatic fragments of α1-CB2 was performed as described previously (Bornstein and Piez, 1966).

Paper Chromatography. In some instances peptide mixtures isolated from column chromatography were further purified by descending chromatography on Whatman No. 3MM paper. The most commonly used solvent system was 1-butanol–acetic acid–water (4:1:1). Samples were applied in streaks and after development, the peptides were located by staining a guiding strip of the paper with the cadmium ninhydrin reagent (Blackburn, 1965). Peptides were eluted from the paper with 1% acetic acid.

NH_2-Terminal Analyses and Sequential Degradation from the NH_2 Terminus. The NH_2-terminal residues of peptides were identified as their dansyl derivatives. The procedure used was a modification of that described by Gray (1967). Peptide samples (5–50 mμmoles) dissolved in 15 μl of 0.2 M triethylamine acetate buffer (pH 9.0) were placed in 6 × 50 mm test tubes and 15 μl of 0.5% dansyl-Cl (Pierce) in acetone was added. After thorough mixing, the reaction was allowed to proceed for 1 hr at 37°. The samples were then dried *in vacuo* and hydrolyzed in 6 N HCl for 16 hr at 108°. When proline or hydroxyproline was suspected as the NH_2-terminal residue, hydrolysis was limited to 6 hr. Dansyl-amino acids were identified by thin-layer chromatography on silica gel plates (Eastman Kodak). The most commonly used solvent systems were benzene–pyridine–acetic acid (16:4:1), chloroform–benzyl alcohol–acetic acid (60:30:3) (Deyl and Rosmus, 1965), and chloroform–*t*-amyl alcohol–formic acid (70:30:1) (Morse and Horecker, 1966). Since the peptides were small and of known amino acid composition, unequivocal identification was usually possible after chromatography in a single solvent system.

Sequential degradation using the Edman reagent was performed as described by Gray (1967). The new NH_2-terminal residue released after each step of degradation was identified as a dansyl derivative. In some instances the resulting PTH-amino acids were also identified. The latter had the advantage of differentiating aspartic acid and glutamic acid residues from their amide forms. The methods used were modification of that described by Edman and Sjöquist (1956) and were previously presented (Piez *et al.*, 1966; Kang *et al.*, 1967).

Amino Acid Analysis. Amino acid analyses were performed on a single-column automatic amino acid analyzer (Beckman) modified for high-speed analysis (Miller and Piez, 1966). Peptides were hydrolyzed in constant-boiling HCl at 108° for 24 hr in tubes sealed under nitrogen. Correction factors

[1] Abbreviations used are: LAP, leucine aminopeptidase; PTH, phenylthiohydantoin; PCA, pyrrolidone-5-carboxylic acid; CPA, carboxypeptidase A.

[2] Pyrrolidonyl peptidase used in this study, prepared by the methods of Doolittle and Armentrout (1968) was a generous gift of Miss L. Fowler and Dr. Edgar Haber, Massachusetts General Hospital, Boston, Mass.

TABLE I: Amino Acid Composition of α1-CB1 and its Enzymatic Fragments.[a]

	α1-CB1 (3–19)	α1-CB1 T1 (3–9)	T1-C1 (3–4)	T1-C2 (5–9)	T1-C2-P1 (5–6)	T1-C2-P2 (7–9)	T2 (10–19)	T2-P1 (10–12)	T2-P2 (13–19)
Aspartic acid	1.0	1.0	–	1.0	–	0.9	–	–	–
Serine	1.9	1.0	1.0	–	–	–	0.9	0.9	–
Glutamic acid	1.1	1.1	–	1.0	–	1.0	–	–	–
Proline	2.1	–	–	–	–	–	2.0	–	1.8
Glycine	3.2	1.1	–	1.0	1.0	–	2.0	1.0	1.1
Alanine	2.1	–	–	–	–	–	2.0	1.0	1.0
Valine	2.0	–	–	–	–	–	2.0	–	1.9
Tyrosine	1.9	1.9	0.9	0.8	0.9	–	–	–	–
Lysine	0.9	0.9	–	1.0	–	1.0	–	–	–
Homoserine[b]	1.0	–	–	–	–	–	1.1	–	1.0

[a] Composition expressed as residue per peptide. A dash indicates 0.1 residue or less, T, C, and P are tryptic, chromotrypic, and peptic fragments, respectively. The numbers in parentheses refer to the residue numbers in Figure 7. [b] Includes homoserine lactone.

for loss of labile amino acids (serine and tyrosine) and incomplete release of valine were used as previously determined (Piez *et al.*, 1960).

Results

Since α1-CB0, α1-CB1, and α1-CB2 constitute the NH_2-terminal sequence in that order, the residue number has been assigned consecutively from the NH_2-terminal residue of α1-CB0. The order of these peptides is based on inferred homology with the α1 chains of rat skin and tendon collagen (Piez *et al.*, 1969; Kang *et al.*, 1969a). Our unpublished data relating to the location of the CNBr peptides of chick skin collagen α1 are also in agreement with this (A. H. Kang and J. Gross, unpublished data). The proposed sequence is presented in Figure 7.

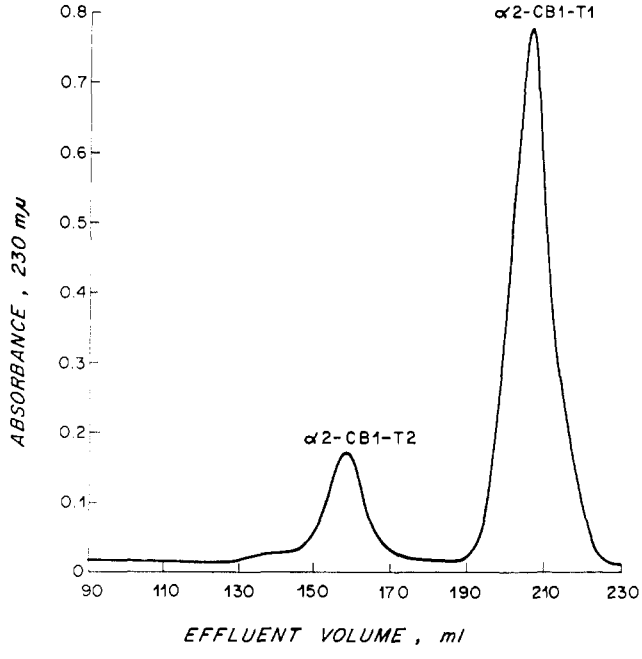

FIGURE 1: Bio-Gel P-2 chromatogram of a tryptic digest of α1-CB1. The enzyme was apparently not eluted from the column under the conditions of the experiment.

The Amino Acid Sequence of α1-CB0 (Residues 1–2). PCA–Hse. This dipeptide gave no detectable NH_2-terminal residue by dansylation. Upon acid hydrolysis it gave rise to one residue each of glutamic acid and homoserine. Since α1-CB0 is a CNBr peptide, homoserine must be the COOH terminus. Therefore, the possibility of an NH_2-terminal PCA was considered. When this peptide was digested with pyrrolidonyl peptidase, subsequent paper chromatography of the reaction mixture indicated the release of PCA and homoserine. Since the cyclization may have occurred during the preparation of the peptide, the presence of the NH_2-terminal PCA *in vivo* cannot necessarily be assumed.

The Amino Acid Sequence of α1-CB1 (Residues 3–19). The tryptic peptides of α1-CB1 could be conveniently separated by column chromatography on Bio-Gel P-2. The elution pattern is presented in Figure 1. Their separation on P-2 despite the similar sizes was due to the retardation of T1 presumably owing to the presence of tyrosine. Its apparently larger extinction at 230 mμ can also be explained on the same basis. The amino acid composition of these peptides is given in Table I.

Peptide α1-CB1-T1 (Residues 3–9). Ser-Tyr-Gly-Tyr-Asp-Glu-Lys. The presence of the lysyl residue in this peptide as well as the presence of the homoseryl residue in α1-CB1-T2 indicates that T1 is on the amino side of T2. Similarly, lysine must be the COOH terminus of T1. A chymotrypsin digestion and paper chromatography of T1 resulted in separation of two peptides, T1-C1 and T1-C2 (see Table I). T1-C2 must be the COOH-terminal peptide since it contains the lysyl residue. The NH_2-terminal residue of α1-CB1-T1-C1 is serine. Thus the sequence of the first two residues of α1-CB1 must be Ser-Tyr. Digestion of α1-CB1-T1-C2 with LAP at an enzyme: substrate ratio of 1:1000 for 6 hr gave Gly (0.56) and Tyr (0.18), whereas hydrolysis at the ratio of 1:70 for

TABLE II: Amino Acid Composition of α1-CB2 and Its Enzymatic Fragments.[a]

	α1-CB2	α1-CB2			
	α1-CB2 (20–55)	Cl (20–39)	C1-T1 (20–(25)	Cl-T2 (26–39)	C2 (40–55)
Hydroxyproline	5.8	3.0	–	3.1	3.0
Serine	0.9	–	–	–	0.9
Glutamic acid	4.0	1.1	–	1.1	3.0
Proline	6.1	4.2	2.2	2.2	2.0
Glycine	12	7.3	2.1	5.1	4.7
Alanine	3.1	2.0	1.0	1.0	1.0
Leucine	1.0	1.0	–	1.0	–
Phenylalanine	1.0	1.0	–	1.0	–
Arginine	1.0	1.0	0.9	–	–
Homoserine[b]	1.1	–	–	–	1.1

[a] Composition expressed as residue per peptide. A dash indicates 0.1 residue or less. T and C are tryptic and chymotryptic fragments, respectively. The numbers in the parentheses refer to the residue numbers in Figure 7. [b] Includes homoserine lactone.

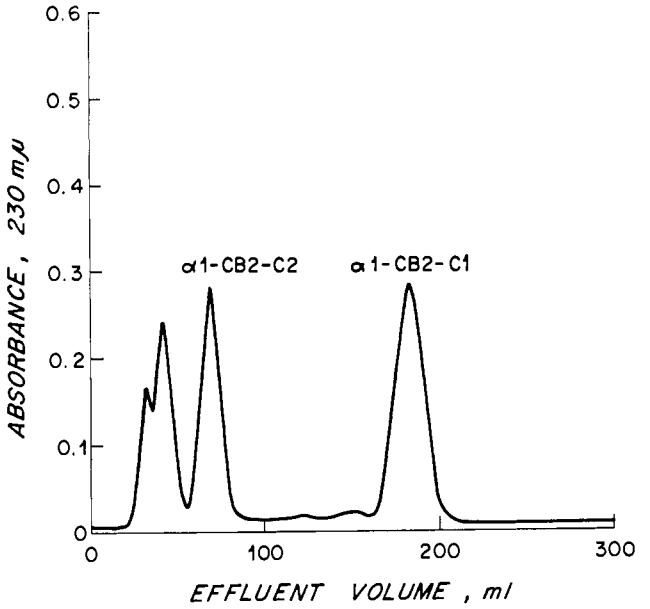

FIGURE 2: Phosphocellulose chromatogram of a chymotryptic digest of α1-CB2. Chromatography was performed at 40° in 0.001 M sodium formate buffer (pH 3.6). A linear gradient of NaCl from 0 to 0.3 M over a total volume of 800 ml was used.

24 hr gave Gly (1.05), Tyr (1.01), Asp (0.35), Glu (0.35), and Lys (0.37). Thus the sequence of α1-CB1-T1-C2 could be deduced to be Gly-Tyr-(Asp, Glu)-Lys. From a peptic digest of T1-C2, two peptide fragments were obtained by paper chromatography (see Table I). Dansylation of T1-C2-P2 showed aspartic acid to be the NH_2 terminus, establishing the sequence of α1-CB1-T1 to be Ser-Tyr-Gly-Tyr-Asp-Glu-Lys.

It is of interest to note that the tyrosyl bond following the glycyl residue is resistant to chymotryptic attack, although susceptible to peptic hydrolysis. Similar resistance of the homologous tyrosyl bond in rat skin collagen to chymotryptic hydrolysis was also noted (Kang *et al.*, 1967).

Peptide α1-CB1-T2 (Residues 10–19). Ser-Ala-Gly-Val-Ala-Val-Pro-Gly-Pro-Hse. A peptic hydrolysis gave rise to two fragments which could be well separated by paper chromatography. Amino acid composition of these fragments is given in Table I. α1-CB1-T2-P2 must be COOH terminal since it has the homoseryl residue. Edman degradation of T2-P1 established the sequence to be Ser-Ala-Gly. The sequence of T2-P2 was also obtained by Edman degradation. Partial confirmation of the sequence was obtained from the experiments with LAP hydrolysis. Digestion of T2-P2 with LAP at an enzyme:substrate ratio of 1:1000 for 1 hr gave Val (1.00) and Ala (0.50), and at the ratio of 1:100 for 24 hr, Val (1.00) and Ala (1.05). No amino acid other than valine and alanine could be liberated by LAP. This limited activity of LAP may be accounted for by the presence of a penultimate prolyl residue. Bacterial collagenase had no activity toward this peptide fragment.

The Amino Acid Sequence of α1-CB2 (Residues 20–55). The first step in the study of α1-CB2 sequence consisted of hydrolysis with chymotrypsin. The chymotryptic peptides were separated by phosphocellulose chromatography. The elution pattern is presented in Figure 2. The first peak eluting at the void volume of the column consisted of nonprotein ultraviolet-absorbing material. The enzyme was apparently not eluted from the column. Since α1-CB2-C1 contains a residue of phenylalanine and α1-CB2-C2 homoserine (see Table II), C1 must be on the amino side of C2. α1-CB2-C1 contained a residue of arginine and was digested with trypsin. The tryptic digest of C1 was then chromatographed on Bio-Gel P-2 (Figure 3). Two well-separated peaks were observed.

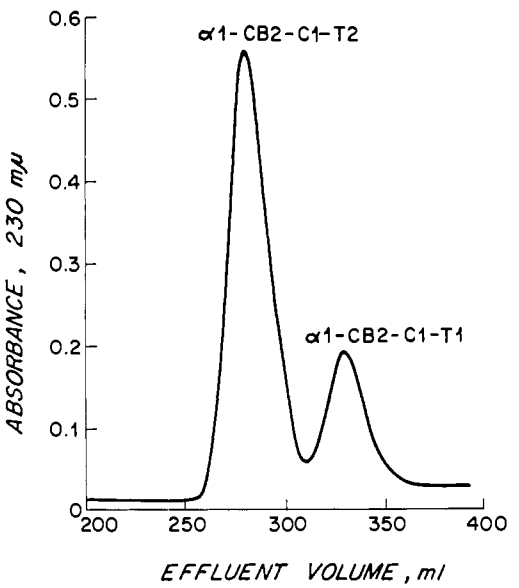

FIGURE 3: Bio-Gel P-2 chromatogram of a tryptic digest of α1-CB2-C1. See text for conditions.

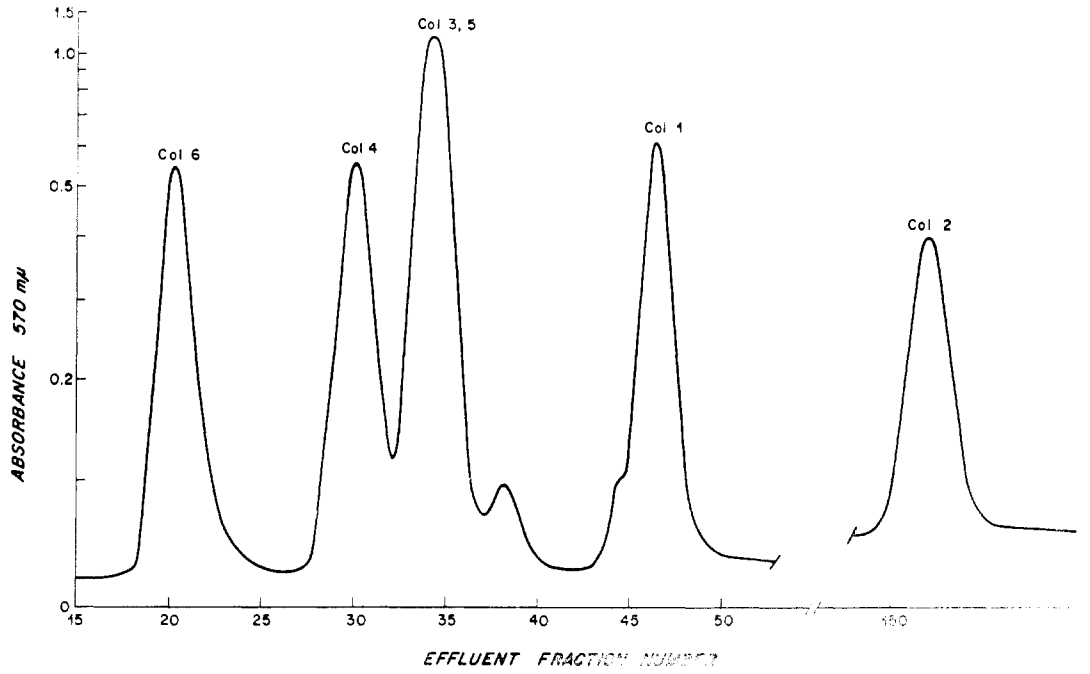

FIGURE 4: Spherix chromatogram of a collagenase digest of α1-CB2-C1. Conditions of chromatography are described in the text.

TABLE III: Amino Acid Composition of the Collagenase Fragments of α1-CB2-C1.[a]

	α1-CB2-C1					
	Col 1 (20–22)	Col 2 (23–25)	Col 3 (26–28)	Col 4 (29–31)	Col 5 (32–34)	Col 6 (35–39)
Hydroxyproline	–	–	1.0	1.0	0.9	–
Glutamic acid	–	–	–	–	–	1.1
Proline	0.9	1.0	–	1.1	–	1.0
Glycine	1.1	1.0	1.0	1.0	1.0	2.1
Alanine	1.0	–	–	–	1.0	–
Leucine	–	–	0.9	–	–	–
Phenylalanine	–	–	–	–	–	1.0
Arginine	–	1.0	–	–	–	–

[a] Composition expressed as residue per peptide. A dash indicates 0.1 residue or less. The numbers in the parentheses refer to the residue numbers in Figure 7.

Apparently, the enzyme was not eluted from the column under the conditions used. α1-CB2-C1-T1 contained a residue of arginine (see Table II) and on the basis of the specificity of trypsin, the arginyl residue was located at position 25.

Peptide α1-CB2-C1 (Residues 20–39). Gly-Pro-Ala-Gly-Pro-Arg-Gly-Leu-Hyp-Gly-Pro-Hyp-Gly-Ala-Hyp-Gly-Pro-Gln-Gly-Phe. The sequence of this region was obtained by the study of the peptide fragments derived by bacterial collagenase digestion. Papain digestion was used to obtain overlapping pieces. The bacterial collagenase digest of α1-CB2-C1 was fractionated on Spherix. Figure 4 depicts the elution pattern. Five well-separated peaks were observed. On paper chromatography in 1-butanol–acetic acid–water, the third peak was resolved into two peptides (α1-CB2-C1-Col 3 and α1-CB2-C1-Col 5). Amino acid composition of these collagenase peptides is given in Table III.

Residues 20–22. Gly-Pro-Ala. Knowing the amino acid composition of α1-CB2-C1-T1 (residues 20–25) an inspection of Table III would make it evident that only the triplet Gly-Pro-Ala meets the requirement of the NH_2-terminal triplet. Edman degradation established the sequence to be Gly-Pro-Ala.

Residues 23–25. Gly-Pro-Arg. The known specificity of trypsin locates arginine at position 25. Edman degradation showed step 1, DNS-glycine; step 2, DNS-proline.

Residues 26–28. Gly-Leu-Hyp. This triplet peptide was separated from the peptide Col 5 by paper chromatography of the fraction obtained from chromatography on Spherix (see Figure 4). The Spherix peak was resolved only into two

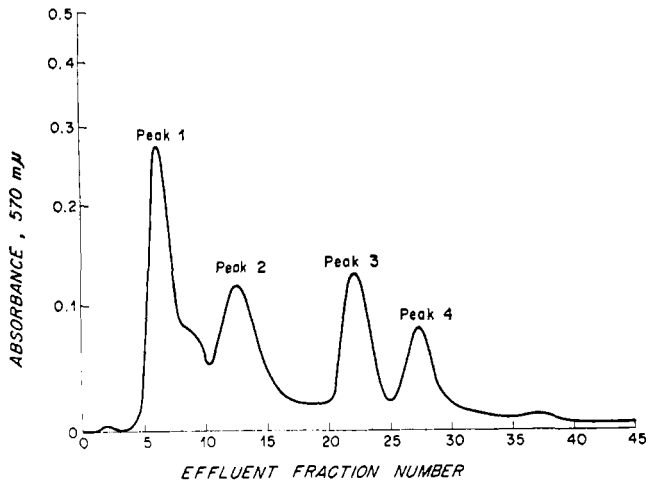

FIGURE 5: Spherix chromatogram of a papain digest of α1-CB2-C2. Conditions of chromatography are described in the text.

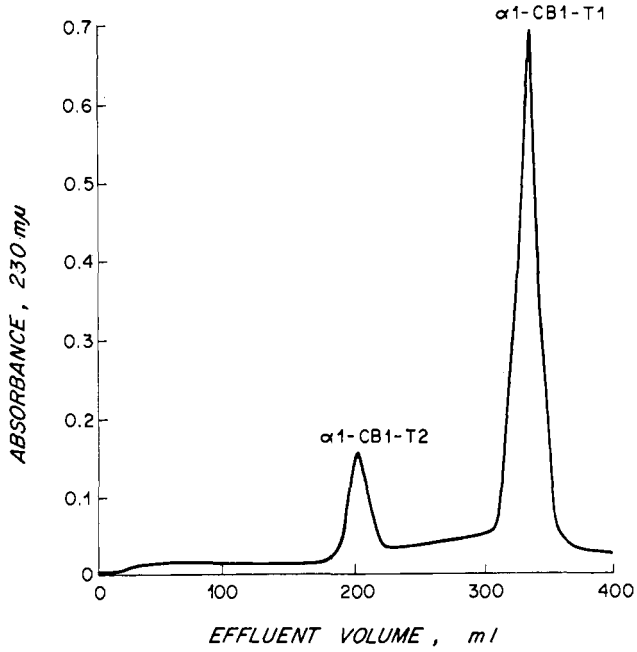

FIGURE 6: Bio-Gel P-2 chromatogram of a tryptic digest of α2-CB1. See text for conditions.

spots; no other components were detectable. Specifically, no evidence was obtained for the possible presence of Gly-Leu-Pro or Gly-Ala-Pro. Edman degradation of this triplet peptide established the sequence to be Gly-Leu-Hyp.

The location of this peptide as residues 26–28, as well as the subsequent triplet peptides, Gly-Pro-Hyp (residues 29–31) and Gly-Ala-Hyp (residues 32–34) was deduced from studies on papain digest of α1-CB2-C1T2 (residues 26–39). When the papain digest was chromatographed on Spherix, five peaks were obtained. Of these, only the material present in the peak eluting at the beginning of the gradient was informative. Paper chromatography in 1-butanol–acetic acid–water (4:1:1) resolved it into two well-separated spots. The peptide fragment in spot 1 (R_F value 0.20) had an amino acid composition of Hyp (1.01), Glu (1.01), Pro (1.05), Gly (1.95), and Ala (0.94). Edman degradation of this peptide showed: step 1, DNS-alanine; step 2, DNS-hydroxyproline. The peptide fragment in spot 2 (R_F value 0.40) had the composition of Hyp (1.92), Pro (1.04), Gly (2.08), and Leu (0.95). DNS-leucine was found by dansylation. Since the glutaminyl residue was found in the last five residues of α1-CB2-C1 (see below), the relative position of the papain fragments and the partial sequence were deduced to be Leu-Hyp-(Gly$_2$, Pro, Hyp)-Ala-Hyp-(Gly$_2$, Pro, Glx).

Residues 29–31. Gly-Pro-Hyp. This tripeptide as obtained from Spherix chromatography (Figure 4) was shown to be pure on paper chromatography. Edman degradation showed step 1, DNS-glycine; step 2, DNS-proline. The location of this tripeptide as residues 29–31 was deduced from analysis of papain fragments as presented above.

Residues 32–34. Gly-Ala-Hyp. This tripeptide was obtained by paper chromatography of the Spherix peak (Figure 4). Edman degradation showed step 1, DNS-glycine; step 2, DNS-alanine. Analysis of papain fragments located the position of this peptide (see above).

Residues 35–39. Gly-Pro-Gln-Gly-Phe. This peptide as obtained from Spherix chromatography was pure. Edman degradation showed step 1, DNS-glycine; step 2, DNS-proline. After step 2, incubation with LAP at an enzyme: substrate ratio of 1:500 for 0.5 hr gave Gln (0.80), Gly (0.21), and Phe (0.20). Hydrolysis of the original peptide fragment with CPA (50 μg of CPA: 0.1 μM peptide) for 1 hr gave Phe (0.90), and Gly (0.24). No other amino acid could be released possibly due to the presence of the penultimate prolyl residue after the glycyl residue.

Peptide α1-CB2-C2 (Residues 40–55). Gln-Gly-Pro-Hyp-Gly-Glu-Hyp-Gly-Glu-Hyp-Gly-Ala-Ser-Gly-Pro-Hse. The sequence of this segment was determined from studies on papain-digested peptide fragments and Edman degradation. Edman degradation of the intact peptide showed step 1, DNS-glutamic acid (glutamine, see below); step 2, DNS-glycine; step 3, DNS-proline; step 4, DNS-hydroxyproline. Digestion with LAP released only glutamine and no other amino acid could be released, probably because the presence of the penultimate prolyl residue prevented release of further amino acids. Despite the presence of an NH_2-terminal glutaminyl residue, apparently a spontaneous cyclization to PCA did not occur under the conditions of the investigation.

A papain digest of α1-CB2-C2 was fractionated on Spherix. Figure 5 represents the elution pattern. Peak 1 was unretarded and was apparently a complex mixture of peptide fragments. Study of the materials in peaks 2, 3, and 4 was sufficient to establish the sequence.

Peak 2 had an amino acid composition of Glu (1.05), Hyp (0.95), and Gly (0.98). Edman degradation showed step 1, DNS-glutamic acid; step 2, DNS-hydroxyproline; step 3, DNS-glycine; step 4, DNS-glutamic acid; step 5, DNS-hydroxyproline. Glutamic acid residues at steps 1 and 4 were identified by the PTH derivative as well.

Peak 3 had amino acid composition of Ala (0.93), Ser (0.95), Gly (1.05), Pro (0.94), and Hse (1.00). Edman degradation showed step 1, DNS-alanine; step 2, DNS-serine; step 3, DNS-glycine. Since homoserine must be the COOH terminus, proline must follow the glycyl residue.

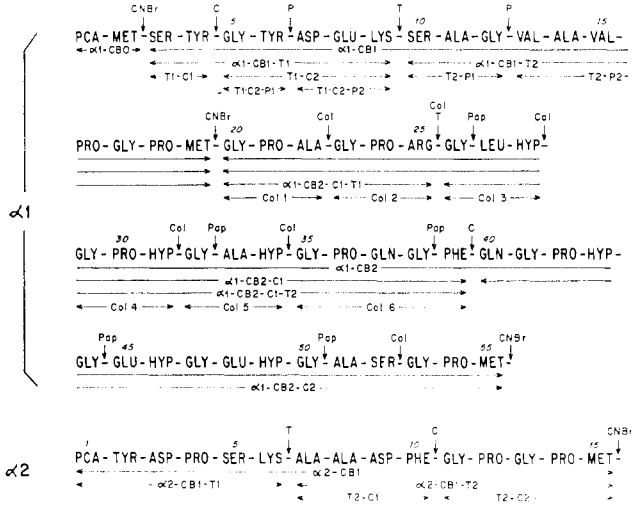

FIGURE 7: Proposed amino acid sequence of the NH_2-terminal peptides of the αl chain (αl-CB0, αl-CBl, and αl-CB2 in that order) and the α2 chain (α2-CBl), and the positions of cleavage by: T, trypsin; C, α-chymotrypsin; P, pepsin; Col, bacterial collagenase; Pap, papain; CNBr, cyanogen bromide. Only the points of major cleavage are indicated for papain.

TABLE IV: Amino Acid Composition of α2-CB1 and Its Enzymatic Fragments.[a]

	α2-CB1				
	α2-CB1 (1–15)	T1 (1–6)	T2 (7–15)	T2-Cl (7–10)	T2-C2 (11–15)
Aspartic acid	2.1	1.1	1.0	1.0	–
Serine	1.0	1.0	–	–	–
Glutamic acid	1.0	1.0	–	–	–
Proline	3.1	1.0	1.8	–	2.1
Glycine	2.0	–	2.1	–	2.0
Alanine	2.0	–	1.9	2.1	–
Tyrosine	0.9	0.9	–	–	–
Phenylalanine	1.0	–	1.0	1.0	–
Lysine	1.0	0.9	–	–	–
Homoserine[b]	1.1	–	1.0	–	1.0

[a] Composition expressed as residue per peptide. A dash indicates 0.1 residue or less. T and C are tryptic and chymotryptic fragments, respectively. The numbers in the parentheses refer to the residue numbers in Figure 7. [b] Includes homoserine lactone.

Amino acid analysis of peak 4 gave Glu (1.00), Gly (2.10), Pro (1.05), and Hyp (0.96). LAP hydrolysis released only glutamine. Edman degradation showed step 1, DNS-glutamic acid; step 2, DNS-glycine; step 3, DNS-proline; step 4, DNS-hydroxyproline.

The data presented here allow the sequence of αl-CB2-C2 to be deduced as presented in Figure 7.

The Amino Acid Sequence of α2-CBl (Residues 1–15). α2-CBl is inferred to be the NH_2-terminal peptide of the α2 chain by homology with the α2 chain of rat skin collagen (Bornstein *et al.*, 1966). Our unpublished data are also in agreement with this conclusion (A. H. Kang and J. Gross, unpublished data). Figure 6 illustrates the elution pattern of tryptic peptides of α2-CBl from Bio-Gel P-2. The presence of a tyrosyl residue in α2-CBl-T1 probably explains the good separation achieved from α2-CBl-T2. T2 contained a residue of phenylalanine and was treated with chymotrypsin. The chymotryptic peptides of T2 were fractionated on PA-35 (Beckman). Two well-separated peaks were obtained. Amino acid composition of these peptide fragments is presented in Table IV.

Peptide α2-CBl-Tl (Residues 1–6). PCA-Tyr-Asp-Pro-Ser-Lys. Since the lysyl residue is located in T1 and the homoseryl residue in T2, T1 must be on the NH_2-side of T2. Upon dansylation no NH_2-terminal amino acid could be detected. Therefore, an NH_2-terminal PCA was considered. After incubation of T1 with pyrrolidonyl peptidase, paper chromatography in 1-butanol–acetic acid–water (4:1:1) separated PCA from a pentapeptide. The latter peptide had an amino acid composition identical with α2-CBl-Tl except for the absence of glutamic acid. Edman degradation of the pentapeptide: step 1, DNS-tyrosine; step 2, DNS-aspartic acid; step 3, DNS-proline. Since lysine must be the COOH terminus, serine is restricted to position 5. The residue 3 was identified as aspartic acid rather than its amide by identification of its PTH derivative.

The tyrosyl bond in the peptide was not susceptible to either chymotryptic or peptic hydrolysis. The ultraviolet absorption spectrum of this peptide was characteristic of normal tyrosine with a peak near 275 mμ in 0.1 N HCl and a shift to a higher wave length in 0.1 N NaOH. Its apparent resistance, at least under the conditions of the present investigation, is unexplained.

Peptide α2-CBl-T2 (Residues 7–15). Ala-Ala-Asp-Phe-Gly-Pro-Gly-Pro-Hse. The sequence of this region was established by the study of chymotryptic peptides. Since α2-CBl-T2-Cl contains a residue of phenylalanine and α2-CBl-T2-C2 homoserine, Cl must be on the amino side of C2 (see Table IV).

Residues 7–10. Ala-Ala-Asp-Phe. This peptide fragment as obtained from PA-35 chromatography was pure. Hydrolysis with LAP at an enzyme:substrate ratio of 1:100 at 37° for 30 min released Ala (1.70), Asp (0.44), and Phe (0.40). No asparagine was detected. Edman degradation showed step 1, DNS-alanine; step 2, DNS-alanine; step 3, DNS-aspartic acid.

Residues 11–15. Gly-Pro-Gly-Pro-Hse. The peptide as obtained from PA-35 chromatography was pure. Edman degradation showed step 1, DNS-glycine; step 2, DNS-proline; step 3, DNS-glycine. Since homoserine must be the COOH terminus, the position of the remaining proline is restricted to position 14.

Discussion

We have presented data establishing the amino acid sequence of three consecutive CNBr peptides from the NH_2 terminus of the αl chain and the NH_2-terminal peptide of the α2 chain of chick skin collagen. The proposed sequence is summarized in Figure 7 along with the points of cleavage

by various enzymes. Only the points of major cleavage are shown in the case of papain.

These data provide an opportunity to compare sequences from a homologous region of collagens from the different animals, chick skin collagen and rat skin collagen. Thus, the sequence of α1-CB1 and α1-CB2 from the two collagens are closely homologous except for a substitution of alanine (chick skin) for serine (rat skin) at position 14 and 20 (Kang *et al.*, 1967; Bornstein, 1967). In addition, the first four residues, PCA-Met-Ser-Tyr, are absent in rat skin α1 although they are present in rat tail tendon α1 (Bornstein, 1969). The biologic significance of this is not known at present. It is possible that the absence of the initial sequences in rat skin collagen may reflect the presence of *in vivo* physiologic proteolytic enzyme. Alternatively, it may be an artifact of isolation of the protein. The heterogeneity in hydroxylation of certain prolyl residues in α1-CB2 of rat skin collagen reported by Bornstein (1967) was not observed in chick skin α1-CB2.

In contrast, there are more marked differences between the amino acid sequence of α2-CB1 of chick skin and that of rat skin collagens (Kang *et al.*, 1967; Bornstein and Kang, 1969) (see Figure 8). The precursor lysyl residue for the cross-link is at position 6 in chick skin collagen, whereas it is at position 5 in rat skin collagen. There is an additional prolyl residue present in the first portion of α2-CB1 of chick skin. Not counting the additional prolyl residue, there are differences in the sequences at 6 positions between the two peptides. Despite these differences there are some important general similarities such as the presence of tyrosyl residues, NH_2-terminal PCA residues, and the precursor lysyl residues for the cross-link within the first several residues.

The relatively strict homology of the sequence of the NH_2-terminal region of α1 and the relative variation in the sequence of the NH_2-terminal region of α2 suggest that the NH_2-terminal region of the α2 chain may be of significance in evolution. Although the sequence of the main body of the collagen molecule is as yet not known, recent studies of the CNBr peptides from rat skin and chick skin collagens (Kang *et al.*, 1969a,b; Butler *et al.*, 1967; Fietzek and Piez, 1969) indicate that the amino acid compositions of the other peptides are closely similar. If these indications are correct, it is possible that the sequence of the NH_2-terminal region of the α2 chain may reflect the evolutionary changes which render species-specific properties to collagen. This concept is supported by the observations that although much of the antigenicity of collagen resides in α1-CB1 and α2-CB1, only α2-CB1 is species-specific immunologically (Michaeli, 1969).

The sequence of the NH_2-terminal region of collagen is of further interest in view of the formation of the intramolecular cross-link in this region (Bornstein *et al.*, 1966; Kang *et al.*, 1969c,d), a process which is of great physiologic importance. Evidence was previously presented (Bornstein *et al.*, 1966; Bornstein and Piez, 1966; Kang *et al.*, 1969c) that the formation of the intramolecular cross-link involves oxidation of ϵ-carbons of lysyl residue 9 of the α1 chain and lysyl residue 6 of the α2 chain to aldehydic groups and that these lysine-derived aldehydes condense in an aldol condensation to form the intramolecular cross-link (Kang *et al.*, 1969d). There is evidence (A. H. Kang and J. Gross, unpublished data) that the intramolecular cross-link, *i.e.*, the aldol condensation product of the two lysine-derived aldehydes, undergoes further reaction to participate in the formation

CHICKEN SKIN	PCA-	TYR-	ASP-	PRO-	SER-	LYS-	ALA-	ALA-	ASP-	PHE-	GLY-	PRO-	GLY-	PRO-	MET-
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
RAT SKIN	PCA-	TYR-	—	-	SER-	ASP-	LYS-	GLY-	VAL-	SER-	ALA-	GLY-	PRO-	GLY-	PRO-	MET-

FIGURE 8: Comparative sequence of α2-CB1 from chick skin and rat skin collagens. The sequence was so arranged to maximize the homology.

of at least one of the intermolecular cross-links. Since these changes probably occur in the extracellular space, the region around the lysyl residues must be accessible to the enzyme or enzymes involved. From the amino acid sequences presented, it is clear that α1-CB1 and α2-CB1 could not exist in the helical form characteristic of the main body of the native collagen molecule since glycine is not every third residue. This unusual structure is also consistent with the previously observed susceptibility of this region to various proteolytic attack (Rubin *et al.*, 1963, 1965; Drake *et al.*, 1966) and CNBr (Bornstein *et al.*, 1966).

Since it is only for rat and chick skin collagens that the amino acid sequence has been studied in a systematic manner, it is not yet possible to evaluate the possible relationship between the primary structure and properties. Nevertheless, it may be anticipated that further accumulation of this kind of information may lead to a better understanding of such a relationship. Furthermore, although no information is available as yet concerning the structure of collagen in human pathology, it is possible that at least in some of the hereditary forms of the diseases involving connective tissue, aberrations in the primary structure of collagen at critical loci may play a pathogenetic role.

Acknowledgments

The authors wish to thank Mrs. M. Bunnell and Mrs. J. Hazard for their skillful technical assistance.

References

Blackburn, S. (1965), *Methods Biochem. Anal. 13*, 1.
Bornstein, P. (1967), *Biochemistry 6*, 3802.
Bornstein, P. (1969), *Biochemistry 8*, 63.
Bornstein, P., and Kang, A. H. (1969), *NATO Conf. Connective Tissue, Santa Marguarita, Italy*.
Bornstein, P., Kang, A. H., and Piez, K. A. (1966), *Biochemistry 5*, 3803.
Bornstein, P., and Piez, K. A. (1966), *Biochemistry 5*, 3460.
Butler, W. T., Piez, K. A., and Bornstein, P. (1967), *Biochemistry 6*, 3771.
Deyl, Z., and Rosmus, J. (1965), *J. Chromatog. 20*, 514.
Doolittle, R. F., and Armentrout, R. W. (1968), *Biochemistry 7*, 516.
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.
Fietzek, P., and Piez, K. A. (1969), *Biochemistry 8*, 2129.
Gray, W. R. (1967), *Methods Enzymol. 11*, 139.
Kang, A. H., Bornstein, P., and Piez, K. A. (1967), *Biochemistry 6*, 788.

Kang, A. H., Faris, B., and Franzblau, C. (1969d), *Biochem. Biophys. Res. Commun. 36*, 345.
Kang, A. H., Igarashi, S., and Gross, J. (1969b), *Biochemistry 8*, 3200.
Kang, A. H., Piez, K. A., and Gross, J. (1969c), *Biochemistry 8*, 3648.
Kang, A. H., Piez, K. A., and Gross, J. (1969a), *Biochemistry 8*, 1506.
Lane, J. M., and Miller, E. J. (1969), *Biochemistry 8*, 2134.
Michaeli, D. (1969), *NATO Conf. Connective Tissue, Santa Marguarita, Italy*.
Miller, E. J., Lane, J. M., and Piez, K. A. (1969), *Biochemistry 8*, 30.
Miller, E. J., and Piez, K. A. (1966), *Anal. Biochem. 16*, 320.
Morse, D., and Horecker, B. L. (1966), *Anal. Biochem. 14*, 429.
Piez, K. A., Bladen, J. A., Bornstein, P., Butler, W. T., Kang, A. H., Lane, J. M., and Miller, E. J. (1969), *Brookhaven Symp. Quant. Biol. 21*, 345.
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.
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.

Hydrogen Ion Titration Curves of the Native Bovine Fibrinogen and of Bovine Fibrinogen Denature of 5 M Guanidine Hydrochloride*

Elemer Mihalyi

ABSTRACT: Titration curves of bovine fibrinogen were obtained at four ionic strengths, at 25°, and at three temperatures at 0.3 ionic strength and also in the presence of 2.02–8.08 M formaldehyde and in 5 M guanidine hydrochloride. Measurements performed in salt solutions, below the isoelectric point, pertain to the denatured molecule, those above this point to the native molecule. Titrations in guanidine were on the denatured molecule. The following stoichiometry gave a good fit to all the data: 119.0 carboxyl (α-, γ-, δ-, and sialic acid carboxyls combined), 15.5 imidazole, 1.2 α-amino, 63.0 ϵ-amino, 29.2 tyrosyl, and 45.0 arginine residues per 10^5 g of fibrinogen. These are in very good agreement with the analytical values. The intrinsic pK's were within their normal range, except that of the guanidino groups, which was low in the native molecule and was normalized in 5 M guanidine hydrochloride. The electrostatic interaction term was smaller in the acid range than in the alkaline one, suggesting an expansion of the molecule on acid denaturation. The isoionic point was found at pH 6.6, more than one unit above the assumed isoelectric point. This suggests the binding of approximately eleven chloride ions to the isoionic protein.

The alkaline section of the titration curves of the native protein can be fitted equally well with large variations in the values of the assumed parameters. The ambiguity was resolved only by the use of all the data from the various types of titrations mentioned above. Simulated titration curves were computed to show the effect of heterogeneity with respect to pK, within one group, on the apparent pK and electrostatic interaction term.

Hydrogen ion titration curves of fibrinogen have been reported by several authors (Nordbö, 1927; Chaudhuri, 1948; Shulman and Ferry, 1950; Mihalyi, 1954). These studies were performed mainly in conjunction with various investigations of the clotting of this protein; therefore, they were restricted to a narrow pH range around neutrality where clotting occurs. Besides the function-oriented motivation of a limited range, an extension of the titration to extreme pH values seemed unprofitable in the absence of adequate knowledge of the behavior of this protein under these conditions. With recent data on the stability and the structural changes of fibrinogen in acid and alkali (Mihalyi, 1965), an interpretation of a complete titration curve was made possible. Therefore, a detailed study of this problem was undertaken and the data obtained are presented in this paper. Titrations were performed at various ionic strengths and temperatures and in the presence of formaldehyde. These were analyzed for the stoichiometry of the various titrated groups and their intrinsic dissociation constants and for electrostatic effects. Titrations were obtained also in the presence of 5 M guanidine hydrochloride. In the latter solvent the protein should approximate a random polypeptide chain, essentially devoid of all the electrostatic and other complicating factors brought about

* From the Section on Cellular Physiology, Laboratory of Biochemistry, National Heart Institute, National Institutes of Health, Bethesda, Maryland 20014, and the Department of Colloid Science, University of Cambridge, Cambridge, England. *Received August 1, 1969.*