Beckman Instruments, Inc., Palo Alto, Calif., Spinco Division.
Vernon, L. P. (1956), *J. Biol. Chem. 222*, 1035.
Weber, K., and Osborn, M. (1969), *J. Biol. Chem. 244*, 4406.
Yphantis, D. A. (1960), *Ann. N. Y. Acad. Sci. 88*, 586.

Chemical Studies on the Cyanogen Bromide Peptides of Rat Skin Collagen. Amino Acid Sequence of α1-CB4*

William T. Butler† and Sharon L. Ponds

ABSTRACT: The amino acid sequence of α1-CB4, cyanogen bromide peptide 3 from the NH_2 terminus of the α1 chain of rat skin collagen, was determined. The peptide contains 47 amino acids arranged in the form $(GlyXY)_{15}GlyHse$ with about four-tenths of the X and Y positions occupied by proline and hydroxyproline. The extent of enzymatic hydroxylation of proline (to form hydroxyproline) was approximately 90% for three residues of the imino acid but was apparently complete in the case of three others. All the prolines acted upon by the hydroxylase were in position Y of the repeating triplet. Two peptide bonds involving lysine and arginine resisted proteolysis by trypsin; these basic amino acids were found to be adjacent to proline and hydroxyproline, respectively. The arginylhydroxyproline peptide bond was slightly susceptible to the enzyme, but no evidence was found to indicate that the lysylproline bond could be cleaved at all. α1-CB4 contains a high proportion of arginine and lysine, relative to that of collagen, an observation consistent with the heavy banding of regions of collagen near its NH_2 terminus when preparations of segment long-spacing aggregates are stained with phosphotungstic acid and viewed by electron microscopy.

Studies from several laboratories have employed CNBr cleavage of α chains of collagen to systematically investigate the chemistry of this important biological molecule (see Butler, 1970a). With this procedure rat skin collagen has been subdivided into 14 CNBr peptides exhibiting differences in composition, chromatographic properties, and presumably amino acid sequence (Butler *et al.*, 1967; Fietzek and Piez, 1969). Eight of the CNBr peptides derive from the two (identical) α1 chains of the molecule and the remaining six derive from the α2 chain. Previous studies on the CNBr peptides have provided information about the amino acid sequence of the NH_2-terminal end of the α1 chain (Kang *et al.*, 1967; Bornstein, 1967; Butler, 1970a,b), the location and identity of intramolecular cross-links (Bornstein and Piez, 1966; Bornstein *et al.*, 1966), the site of the initial cleavage of the collagen molecule by tadpole collagenase (Kang *et al.*, 1966), the location and distribution of the carbohydrate in α1 (Butler, 1970a), the location and probable nature of hydroxylamine-sensitive bonds (Butler, 1969; Bornstein, 1970), and the degree of species variability of collagen (Bornstein and Kang, 1970; Bornstein and Nesse, 1970). Additionally, studies relating to the antigenicity of collagen (Michaeli *et al.*, 1969; Timpl *et al.*, 1970; Bornstein and Nesse, 1970) and the occurrence of a more primitive type (type II) of collagen in cartilage (Miller and Matukas, 1970) have been facilitated by use of CNBr peptides.

In this communication we present data that indicate the amino acid sequence of α1-CB4, CNBr peptide 3 from the NH_2 terminus of the α1 chain of rat skin collagen. This peptide consists of 47 amino acids[1] with relatively high levels of the basic amino acids (arginine and lysine) and of the imino acids (proline and hydroxyproline).

Experimental Section

Preparation of α1-CB4. Lathyritic collagen was prepared (Bornstein and Piez, 1966) and then, without prior fractionation of α chains, cleaved with CNBr. Collagen samples were dissolved in 70% formic acid at a concentration of 10 mg/ml and incubated with a hundredfold excess (relative to methionine) of CNBr at room temperature for 4 hr. The samples were desalted by gel filtration on columns of Bio-Gel P2 (Bio-Rad Laboratories) and lyophilized.

The CNBr peptides from whole collagen were separated by modification of a method used previously (Butler, 1970a). Samples, ranging from 100 to 300 mg, were dissolved in 15 ml of 0.02 M sodium citrate (pH 3.6) and applied to 2.5 × 30 cm columns of CM-cellulose (Whatman CM-32) at 40°. Columns were then eluted with 50 ml of 0.02 M sodium citrate (pH 3.6) and finally with a linear gradient formed from 1100 ml each of 0.02 M sodium citrate (pH 3.6)–0.02 M NaCl and 0.02 M sodium citrate (pH 3.6)–0.14 M NaCl. The effluent was continuously monitored with a Beckman DB-G spectro-

* From the Institute of Dental Research and Department of Biochemistry, the University of Alabama, Birmingham, Alabama 35233. *Received December 9, 1970.* A preliminary report of these investigations has appeared (Butler, 1970b). This research was sponsored by U. S. Health Service Grant DE-02670 from the National Institute of Dental Research.

† To whom to address correspondence.

[1] The original report concerning the CNBr peptides of α1 indicated that α1-CB4 contained 46 amino acids (Butler *et al.*, 1967). Data reported in the present communication indicate that the correct composition includes 16 glycines instead of 15 reported in the earlier paper. The present data are consistent with those reported by Miller *et al.* (1969) and Kang *et al.* (1969) for α1-CB4 from chick collagens.

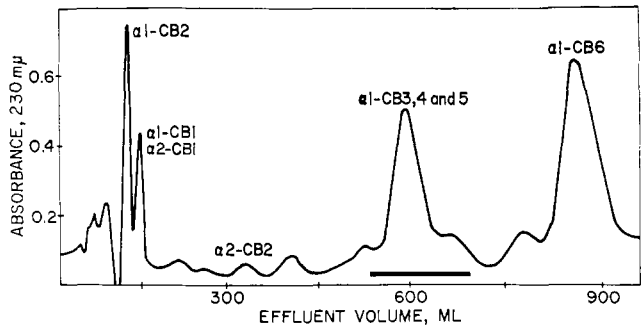

FIGURE 1: Chromatography of 300 mg of CNBr-treated lathyritic rat skin collagen on CM-cellulose at 40°. Elution was with (1) 50 ml of 0.02 M sodium citrate (pH 3.6) and (2) a linear gradient formed from 1100 ml each of 0.02 M sodium citrate (pH 3.6)–0.02 M NaCl and 0.02 M sodium citrate (pH 3.6)–0.14 M NaCl. Fractions of 15 ml were collected.

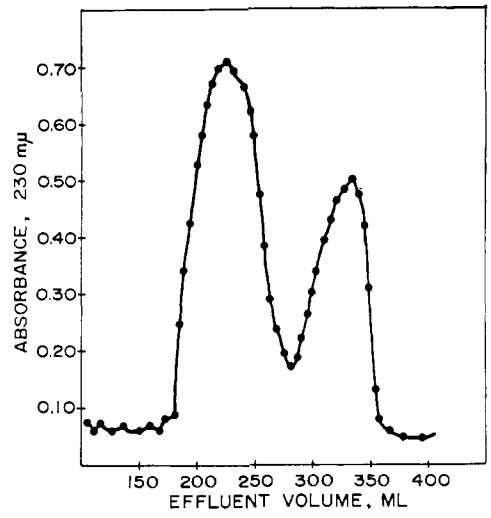

FIGURE 2: Gel filtration on Sephadex G-50 of 50 mg of material (containing α1-CB3, α1-CB4, and α1-CB5) from the fraction of CM-cellulose chromatography marked with a bar (Figure 1). The gel column (1.8 × 130 cm) was equilibrated and eluted with 0.2 M acetic acid.

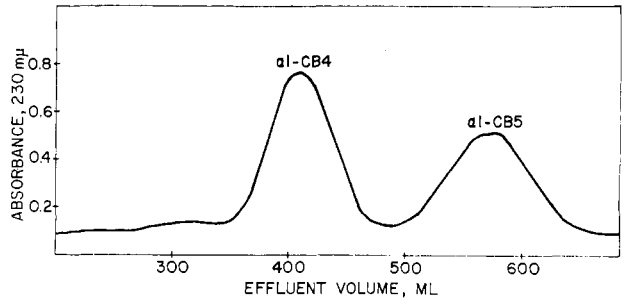

FIGURE 3: Phosphocellulose chromatography of the fraction from filtration on Sephadex G-50 (Figure 2) that contained α1-CB4 and α1-CB5 (the fraction eluted between 290 and 350 ml). The column was eluted with a linear gradient formed from 750 ml each of 0.001 M sodium formate (pH 3.6)–0.1 M NaCl and 0.001 M sodium formate (pH 3.6)–0.8 M NaCl.

photometer equipped with a flow cell. Figure 1 shows the early part of a typical chromatogram developed in this way. The material in the fraction marked by the solid line consisted mainly of α1-CB3, α1-CB4, and α1-CB5; it was desalted on columns of Bio-Gel P2 and further fractionated by gel filtration of Sephadex G-50 (Pharmacia) into two fractions (Figure 2). The larger ultraviolet-absorbing peak, emerging shortly after the void volume, contained principally α1-CB3; the material in the smaller peak contained α1-CB4 and α1-CB5. The material in this smaller peak was lyophilized and further purified by ion-exchange chromatography on phosphocellulose in a manner similar to that described by Kang *et al.* (1969). Phosphocellulose columns, prepared as described by Bornstein and Piez (1966), were equilibrated with 0.001 M sodium formate (pH 3.6) containing 0.1 M NaCl (starting buffer) and then eluted with a linear gradient formed from 750 ml each of the starting buffer and 0.001 M sodium formate (pH 3.6) containing 0.8 M NaCl. As illustrated in Figure 3, two peptide peaks that contained α1-CB4 and α1-CB5 were observed in approximately equimolar quantities. All samples of α1-CB4 prepared in this way had amino acid compositions identical with those of samples prepared by a different procedure (Butler *et al.*, 1967). They also contained serine levels of less than 0.05 residue/peptide; this indicated high purity, because α1-CB4 is the only α1-derived, CNBr peptide that does not contain this amino acid.

Amino Acid Analysis. Peptides were hydrolyzed in constant-boiling HCl at 108° for 16–18 hr in an atmosphere of nitrogen. Analyses were performed on a Beckman 120C automatic amino acid analyzer modified for single-column high-speed analysis (Miller and Piez, 1966). No corrections were made for losses of threonine.

Enzymatic Digestion of Peptides. The methods used for cleavage of peptides with trypsin, collagenase, and pepsin have been described (Butler, 1970a).

Separation of Peptides. After proteolytic cleavage of a peptide, the products were separated by ion-exchange chromatography, paper chromatography, or high-voltage paper electrophoresis as described (Butler, 1970a).

Stepwise (Edman) Degradation. The degradative procedure of Blömback *et al.* (1966) was followed in the derivatization and cyclization of peptides. After cyclization the trifluoroacetic acid was removed with a stream of air, fractions were redissolved in 0.2 ml of 0.2 M acetic acid and the amino acid derivatives were extracted with three portions (0.5 ml) of either ethyl acetate or butyl acetate. Derivatives were then dried and converted into PTH[2]-amino acids by heating for 1 hr at 80°, under nitrogen in 0.2 ml of 0.1 M HCl in 20% ethanol. The resultant PTH-amino acids were then extracted from the aqueous phase with three portions (0.5 ml) of ethyl acetate. Extracts were combined, dried, redissolved in a small volume of acetone, and spotted on thin-layer plates of silica gel (Eastman Organic Chemicals) that contained a fluorescent indicator. Chromatograms were then developed in solvents D and E of Edman and Sjöquist (1956).

The PTH-amino acids, detected by fluorescence quenching under an ultraviolet lamp, were identified by comparing their chromatographic mobilities with those of standards run on each side of the unknown. Using this procedure, it was possible to reliably determine the sequence of as many as 15 of the amino acids in a peptide that contained 20 or more amino acids. The method is not applicable to smaller peptides (containing less than 10 amino acids) since these peptides are usually slightly soluble in organic solvents. The procedure for

[2] Abbreviations used are: PTH, phenylthiohydantoin; PTC, phenylthiocarbamyl.

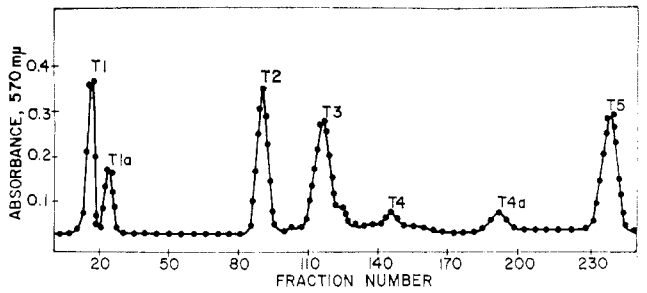

FIGURE 4: Ion-exchange chromatography of the tryptic digest of α1-CB4 on a 0.9 × 150 cm column of Dowex 50-X4. Peptides were eluted from the column with (1) 50 ml of 0.2 M pyridine acetate (pH 3.1) and (2) a linear gradient formed from 750 ml each of 0.2 M pyridine–acetate (pH 3.1) and 2.0 M pyridine–acetate (pH 5.0) (Schroeder, 1967). Effluent fractions of 5 ml were collected and analyzed for ninhydrin-positive material with a Technicon Auto-Analyzer.

subtractive Edman degradation has been described (Butler, 1970a).

Results

Tryptic Peptides from α1-CB4. The mixture of peptides resulting from hydrolysis of α1-CB4 with trypsin was separated into seven ninhydrin-positive peaks by ion-exchange chromatography on Dowex 50 (Figure 4). Amino acid analyses indicated that the compositions of five of the tryptic peptides, T1, T2, T3, T4, and T5, accounted for all the amino acids of α1-CB4 (Table I); thus these five peptides represent the complete amino acid sequence of α1-CB4.

The apparent discrepancy between this low number of peptides and the arginine and lysine content of α1-CB4 (and thus the number of bonds presumably susceptible to proteolysis by trypsin) was clarified by the observation that peptide T4 contained two residues of arginine and one of lysine. Peptide T1a had a composition identical with that of peptide T1. These two peptides probably differed from each other by the occurrence of homoserine as the lactone in T1a and as the open form in T1. Peptide T4a differed from T4 only with respect to four amino acids located at the COOH terminus of the latter peptide. The proof and significance of this observation are discussed in a later section.

The order of the 5 tryptic peptides of α1-CB4 was shown to be T5-T2-T4-T3-T1 by the following data. Edman degradation using the identification of the PTH-amino acids indicated the following partial sequence for α1-CB4: Gly-Pro-Arg-Gly-Pro-Hyp-Gly-Pro-Hyp-Gly-X-Asn-Gly-Asp-Asp. The amino acid compositions (Table I) and sequence data (given in later sections) show that the only possible arrangement of tryptic peptides that would allow this sequence at the NH_2 terminus of α1-CB4 is T5-T2-T4. Though the PTH-amino acid for the position marked X in the above sequence was not identified, the amino acid must be the lysine located at the COOH terminus of T2. α1-CB4 was liberated from the α1 chain of collagen by digestion with CNBr; this treatment converts methionine into homoserine (lactone) at the COOH terminus of the resultant peptides (Gross, 1967). Since T1 contained homoserine but no lysine or arginine, peptide T1 must have been tryptic peptide 5 in the alignment. These data allow the deduction that T3 is tryptic peptide 4.

Amino Acid Sequence of T5 (Residues 1–3). One step of subtractive Edman degradation removed a residue of glycine from this tripeptide. Since it was obtained from a trypsin digest of α1-CB4, the arginine residue of T5 must be in the COOH-terminal position (Hill, 1965). Its sequence is therefore: Gly-Pro-Arg.

Amino Acid Sequence of T2 (Residues 4–11). Three steps of subtractive Edman degradation indicated the partial sequence Gly-Pro-Hyp(Gly$_2$,Pro,Hyp)Lys for T2. The specificity of trypsin indicates that lysine was the COOH-terminal amino acid. To obtain further data concerning its sequence, a sample of T2 was hydrolyzed with collagenase and the products separated by paper electrophoresis at pH 6.5. Two ninhydrin-positive spots of peptide material were observed, one migrating as a positively charged component and the second as neutrally charged. The cationic peptide had the composition:

TABLE I: Amino Acid Analysis of α1-CB4 and of Peptides Resulting from Its Hydrolysis with Trypsin.[a]

Amino Acid	α1-CB4	T1	T1a	T2	T3	T4	T4a	T5
4-Hydroxyproline	5.9	1.9	2.2	1.8	1.0	1.1	—	—
Aspartic acid	3.0	—	—	—	—	2.9	2.9	—
Threonine	1.0	1.0	1.0	—	—	—	—	—
Glutamic acid	3.0	—	—	—	1.0	2.1	1.1	—
Proline	5.8	—	—	2.3	2.0	0.9	0.9	1.0
Glycine	15.8	4.2	4.0	3.1	3.0	4.9	4.1	1.0
Alanine	3.0	1.1	1.0	—	1.0	1.0	1.0	—
Leucine	2.0	2.0	1.9	—	—	—	—	—
Lysine	2.0	—	—	0.9	—	1.0	0.9	—
Arginine	4.1	—	—	—	1.0	2.0	1.0	1.0
Homoserine[b]	1.0	1.0	1.0	—	—	—	—	—
Total	47	11	11	8	9	16	12	3

[a] Compositions are expressed as residues per peptide. A dash indicates either the absence of an amino acid or that the content was less than 0.1 residue/peptide. Values are the average of analyses on two or three separate preparations. [b] Homoserine includes homoserine lactone.

TABLE II: Subtractive Edman Degradation of T4.[a]

Amino Acid	Step 0	1	2	3	4
Hydroxyproline	1.07	0.87	1.04	1.04	1.06
Aspartic acid	2.94	*2.18*	1.85	*1.28*	*0.66*
Glutamic acid	2.12	2.17	2.00	2.05	2.19
Proline	0.90	0.99	0.88	0.81	0.99
Glycine	4.88	4.88	*3.99*	3.85	4.06
Alanine	1.01	1.02	1.04	0.99	0.99
Lysine	0.98	0.86	0.85	0.76	0.72
Arginine	1.95	2.06	2.07	1.83	1.87
Sequence		Asx	Gly	Asx	Asx

[a] The amino acid composition after each step is expressed as residues of amino acid per peptide. The amino acid removed at each step is indicated below the appropriate column. The composition of the original peptide is in the first column (step 0).

TABLE III: Amino Acid Compositions of Peptides Recovered after Hydrolysis of α1-CB4 with Pepsin.[a]

Amino Acid	P1	P2
4-Hydroxyproline	1.7	3.6
Aspartic acid	3.0	0.3
Threonine	—	0.9
Glutamic acid	1.1	2.0
Proline	3.2	3.1
Glycine	6.2	9.6
Alanine	—	2.7
Leucine	—	1.8
Lysine	1.0	1.0
Arginine	1.0	2.9
Homoserine	—	1.0
Total[b]	17	30

[a] Residues per peptide. [b] The total represents the sum of the nearest integers of the amino acids.

glycine, 1.22; and lysine, 0.78; it was not further purified. The material from the second spot was subjected, for further purification, to paper chromatography in butanol–acetic acid–water (12:3:5, v/v). Only one major ninhydrin-positive spot of R_F 0.19 was observed; it gave the amino acid composition: glycine, 0.93; proline, 1.17; hydroxyproline, 0.90. Two steps of Edman degradation sequentially removed residues of glycine and proline, indicating that the sequence was Gly-Pro-Hyp. These data are consistent with those from Edman degradation of α1-CB4 (see section on tryptic peptides from α1-CB4) and show that the amino acid sequence of T2 is Gly-Pro-Hyp-Gly-Pro-Hyp-Gly-Lys.

Sequence of T4 (Residues 12–27). Subtractive Edman degradation (Table II) indicated the sequence Asx-Gly-Asx-Asx at the NH_2 terminus of peptide T4. In a separate experiment, stepwise degradation of peptide T4 yielded the following derivatives: step 1, PTH-asparagine; step 2, PTH-glycine; step 3, PTH-aspartic acid; step 4, PTH-aspartic acid; step 5, PTH-glycine; step 6, PTH-glutamic acid; step 7, PTH-alanine; step 8, PTH-glycine.

Experiments on cleavage of α1-CB4 with pepsin, produced information concerning the sequence of amino acids at the COOH-terminal portion of T4. Pepsin digestion of α1-CB4 yielded only two peptides separated by ion-exchange chromatography on Dowex 50-X4. Amino acid analysis of these two peptides (Table III) indicated that they contained 17 (P1) and 30 (P2) residues, and that all the amino acids of α1-CB4 were recovered. Peptide P2 contained homoserine and was thus from the COOH-terminal portion of α1-CB4. These data indicated that pepsin cleavage took place at the peptide bond involving residues 17 and 18 of α1-CB4. Edman degradation of P2 gave the following PTH-amino acids: step 1, PTH-alanine; step 2, PTH-glycine; step 3, α-PTH-ε-PTC-lysine; step 4, PTH-proline; step 5, PTH-glycine; step 6, PTH-arginine; step 7, PTH-hydroxyproline (plus a trace of PTH-proline); step 8, PTH-glycine; step 9, PTH-glutamic acid. These data indicate the following amino acid sequence for T4: Asn-Gly-Asp-Asp-Gly-Glu-Ala-Gly-Lys-Pro-Gly-Arg-Hyp-Gly-Glu-Arg.

The lysyl and arginyl bonds in T4, which resisted cleavage by trypsin, involved the imino acids proline and hydroxyproline. The arginylhydroxyproline bond is susceptible to cleavage, but apparently at a slow rate since one of the tryptic peptides of α1-CB4, T4a, gave the same amino acid composition (Table I) as that of T4 except for loss of one residue each of hydroxyproline, glycine, glutamic acid, and arginine. The expected tetrapeptide (Hyp-Gly-Glu-Arg) resulting from cleavage of this bond has not been observed. This tetrapeptide would presumably give a very low color yield with ninhydrin since hydroxyproline would be its NH_2-terminal amino acid.

TABLE IV: Subtractive Edman Degradation of Peptide T3.[a]

	Step					
Amino Acid	0	1	2	3	4	5
Hydroxyproline	1.04	0.95	0.87	—	—	—
Glutamic acid	1.05	0.97	1.03	1.03	1.03	1.15
Proline	1.93	2.04	*1.11*	1.07	1.01	*0.30*
Glycine	2.88	*2.08*	1.99	1.99	*1.30*	1.37
Alanine	0.96	1.03	1.00	0.97	0.96	1.01
Arginine	1.03	0.98	1.01	1.01	0.98	0.84
Sequence		Gly	Pro	Hyp	Gly	Pro

[a] Explanations are in Table II.

Amino Acid Sequence of T3 (Residues 28–36). The following data indicated that the probable sequence of peptide T3 was Gly-Pro-Hyp-Gly-Pro-Gln-Gly-Ala-Arg. Results from subtractive Edman degradation (Table IV) gave the partial sequence Gly-Pro-Hyp-Gly-Pro at the NH_2 terminus of T3. Digestion of T3 with collagenase and separation of the products by electrophoresis at pH 6.5 yielded two ninhydrin-positive components with the amino acid compositions given in Table V. One step of Edman degradation of the tripeptide, T3-CL2, removed glycine, showing that its sequence is Gly-Ala-Arg. Arginine was known to be the COOH-terminal amino acid because T3 was isolated after cleavage of α1-CB4 with trypsin (Hill, 1965). The hexapeptide T3-Cl1 migrates as a neutral compound; it must therefore have contained glutamine rather than glutamic acid. By deduction, glutamine was placed in position 6 of peptide T3.

Amino Acid Sequence of T1 (Residues 37–47). Subtractive Edman degradation (Table VI) indicated that the sequence at the NH_2 terminus of peptide T1 was Gly-Leu-Hyp. Ad-

TABLE V: Amino Acid Compositions of Peptides Isolated after Hydrolysis of Peptide T3 with Collagenase.[a]

Amino Acid	T3-CL1	T3-CL2
Hydroxyproline	1.04	—
Glutamic acid	0.93	—
Proline	2.18	—
Glycine	1.89	0.94
Alanine	—	0.96
Arginine	—	1.01
Total[b]	6	3

[a] Values are given as residues per peptide. [b] The total is the sum of the nearest integers of the amino acids.

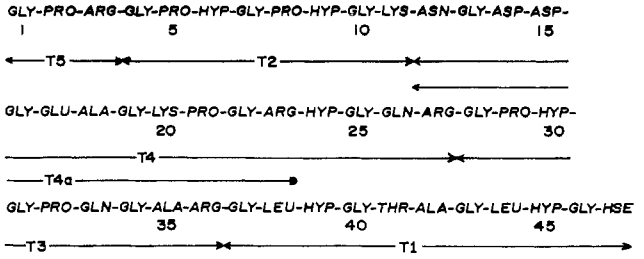

FIGURE 5: The amino acid sequence of α1-CB4. The tryptic peptides discussed in the text are delineated by the arrows.

ditional information concerning the sequence of peptide T1 was obtained by collagenase digestion of α1-CB4, which produced a number of small peptides, that were partially resolved by ion-exchange chromatography on Dowex 50-X4. Amino acid analysis indicated that one of the peptides, CL1, contained six amino acids and was relatively pure. Subtractive Edman degradation of this component (Table VII) indicated that its partial amino acid sequence was Gly-Thr-Ala-Gly-(Leu,Hyp). Since the only threonine and leucine of α1-CB4 are found in peptide T1, and since its sequence is incompatible with its location at the NH_2 terminus of T1, peptide CL1 must have originated from the sequence represented by residues 40–45 of α1-CB4.

To obtain information about the amino acid sequence at the COOH-terminal portion of peptide T1, 0.5 μmole of the peptide was hydrolyzed for 20 min in constant-boiling HCl at 108° (Light, 1967). The resultant small peptides from this partial acid hydrolysate were separated by descending paper chromatography in butanol–acetic acid–water (12:3:5, v/v). Seven major ninhydrin-positive spots were observed. Peptides from the appropriate areas of the chromatogram were eluted from the paper with 2% acetic acid and subjected to amino acid analysis. One peptide obtained in this manner displayed an R_F of 0.25 and gave the composition: hydroxyproline, 1.08; glycine, 1.09; homoserine, 0.82. One step of subtractive Edman degradation on this tripeptide removed a residue of hydroxyproline. Since homoserine occupies the COOH-terminal position of α1-CB4 and of T1 (see above), the amino acid sequence of this tripeptide was Hyp-Gly-Hse. By deduction, the remaining leucine of peptide T1 can be placed at position 8. These data indicate that the amino acid sequence of peptide T1 is Gly-Leu-Hyp-Gly-Thr-Ala-Gly-Leu-Hyp-Gly-Hse.

Extent of Hydroxylation of Proline in α1-CB4. Edman degradation yielded a trace of PTH-proline for steps corresponding to positions 6, 9, and 24 (step 7 of peptide P2) of α1-CB4. The amino acid analysis of peptides T2 and P1 (Tables I and III) consistently gave values for hydroxyproline that were lower than integrals while the proline values were slightly higher than integrals. These observations suggest that the enzymatic hydroxylation of prolyl to form the hydroxyprolyl residues at three positions of α1-CB4 is incomplete. On the other hand, amino acids in positions 30, 39, and 45 appear to be entirely in the form of hydroxyproline. The values for proline and hydroxyproline in peptide T3 were always quite close to integrals (Table I). After step 3 of Edman degradation of peptide T3, the loss of a full residue of hydroxyproline was observed without the loss of proline (Table IV). No proline was found in peptide T1 (Tables I and VI), thus indicating that the imino acids at positions 39 and 45 of α1-CB4 were fully hydroxylated. These conclusions should be considered tentative until more definitive experiments have been performed.

TABLE VI: Subtractive Edman Degradation of Peptide T1.[a]

Amino Acid	Step 0	1	2	3
4-Hydroxyproline	1.84	1.84	1.81	*0.86*
Threonine	0.94	0.98	0.92	1.00
Glycine	4.13	*3.10*	2.95	2.86
Alanine	1.04	1.05	0.93	1.03
Leucine	1.91	1.93	*1.02*	1.01
Homoserine	1.13	1.09	1.27	0.97
Sequence		Gly	Leu	Hyp

[a] Explanations are in Table II.

TABLE VII: Subtractive Edman Degradation of Peptide CL1.[a]

Amino Acid	Step 0	1	2	3	4
4-Hydroxyproline	1.05	0.81	0.99	1.00	0.98
Threonine	1.00	1.08	...	...	...
Glycine	2.18	*1.22*	1.25	1.12	*0.68*
Alanine	1.04	0.92	1.02	*0.25*	0.27
Leucine	0.92	1.01	0.99	0.88	1.02
Sequence		Gly	Thr	Ala	Gly

[a] Explanations are in Table II. Nomenclature for and significance of this peptide are discussed in the section on the amino acid sequence of T1.

Discussion

The amino acid sequence of α1-CB4 deduced from the studies reported here is shown in Figure 5. α1-CB4 is CNBr peptide 3 from the NH_2-terminal end of the α1 chain of rat skin collagen and represents residues 52–98 of this polypeptide. The sequence of α1-CB1 (residues 1–15), α1-CB2 (residues 16–51), and α1-CB5 (residues 99–135) have been reported (Kang *et al.*, 1967; Bornstein, 1967; Butler, 1970a,b).

The structure of α1-CB4 displays features postulated to occur in the helical regions of collagen. For example, glycine occurs as every third amino acid and the imino acids, proline and hydroxyproline, occur in more than a third of the remaining positions.

An unusual feature of the portion of α1 from which α1-CB4 and most of α1-CB5 were derived (encompassing residues 52–120 of α1) is the relatively high content of basic amino acids. Thus five arginyl, four lysyl, one histidyl, and one hydroxylysyl residues are located within a span of 70 amino acids, a concentration of basic amino acids approximately twice that in the α1 chain itself. In contrast only one residue each of arginine and lysine is found in the first 51 residues of the α1 chain; the latter is oxidatively deaminated as a first step in the formation of cross-links (Bornstein and Piez, 1966; Bornstein *et al.*, 1966). This same span, bearing the large number of basic side chains, contains only four negatively

charged groups. The area near the A end of segment long-spacing fibrils, which stains heavily with phosphotungstic acid, may correspond to the region of $\alpha 1$ from which $\alpha 1$-CB4 and $\alpha 1$-CB5 originated (see Hodge, 1967).

Grassmann and coworkers noted that collagen contained a number of lysyl and arginyl bonds that were susceptible to cleavage by trypsin (Grimm and Grassmann, 1964; Hannig and Nordwig, 1968). Based on the frequent occurrence of proline and hydroxyproline in collagen, the resistance of peptide bonds involving proline to proteolysis, and experiments with synthetic peptides, these workers postulated that collagen contained trypsin-resistant, peptide bonds involving both arginyl and lysyl residues adjacent to those of the imino acids, proline and hydroxyproline. The studies outlined here offer direct proof for this hypothesis, demonstrating a lysylproline bond (residues 20–21) and an arginylhydroxyproline bond (residues 23–24). It is of interest that the arginylhydroxyproline bond is susceptible to cleavage by trypsin but apparently at a rate substantially lower than for other arginyl peptide bonds. The lysylproline bond appears completely resistant to hydrolysis by trypsin. A similar observation has been made for an arginylhydroxyproline bond in $\alpha 2$-CB2, a small peptide from the central portion of $\alpha 2$ (Highberger *et al.*, 1970).

Bornstein has shown that certain prolyl residues in $\alpha 1$-CB2, which are candidates for enzymatic hydroxylation, are incompletely hydroxylated (Bornstein, 1967). This observation seemed to Prockop (1970) to be quite compatible with studies on the mechanism of hydroxylation of proline. The studies reported here identify three more imino acid residues in collagen (positions 6, 9, and 24 of $\alpha 1$-CB4) that are apparently incompletely hydroxylated. Since $\alpha 1$-CB4 is from the portion of $\alpha 1$ adjacent to the COOH-terminal region of $\alpha 1$-CB2, all residues of this imino acid that are reported to be incompletely hydroxylated are in the same vicinity, near the NH_2 terminus of collagen. In view of these observations it appears possible that proline hydroxylase may have a slightly lowered affinity for this portion of the molecule. This conclusion seems even more valid when one considers that three additional residues located in the COOH-terminal portion of $\alpha 1$-CB4 (at positions 30, 39, and 45) and three in $\alpha 1$-CB5 (Butler, 1970a) appear to be completely hydroxylated. The phenomenon of incomplete hydroxylation of proline is not limited to this region of the molecule but also applies to the region from which $\alpha 1$-CB8 is derived (Bornstein, 1970).

All the residues of hydroxyproline examined in this study were in position three of the collagen triplet (the repeating sequence of the form Gly-X-Y with glycine designated at position 1). The specificity of proline hydroxylase may dictate this phenomenon in higher animals. Recent work has shown the occurrence of sequences from earthworm cuticle collagen containing hydroxyproline in position two of the collagen triplet (Goldstein and Adams, 1970); this may indicate the presence, in lower animals, of hydroxylases with differing specificities. Though the proline in the sequence Gly-Lys-Pro in $\alpha 1$-CB4 (positions 19–21) is in position three, it seems not to be a substrate for proline hydroxylase. The lysine side chain may in some way prevent the hydroxylation; however, the side chain of arginine does not interfere with hydroxylation, for the sequence Gly-Arg-Hyp was also noted (positions 22–24 of $\alpha 1$-CB4).

Acknowledgments

We are indebted to Mr. John E. Finch, Jr., and Mr. Roy Pierce for technical assistance. We acknowledge the excellent artistic assistance of Miss Jeanette Mills.

References

Blömback, B., Blömback, M., Edman, P., and Hessel, B. (1966), *Biochim. Biophys. Acta 115*, 371.

Bornstein, P. (1967), *Biochemistry 6*, 3082.

Bornstein, P. (1970), *Biochemistry 9*, 2408.

Bornstein, P., and Kang, A. H. (1970), *in* Chemistry and Molecular Biology of the Intercellular Matrix, Tristram, G. R., and Balazs, E. A., Ed., London, Academic Press, p 99.

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

Bornstein, P., and Nesse, R. (1970), *Arch. Biochem. Biophys. 138*, 443.

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

Butler, W. T. (1969), *J. Biol. Chem. 244*, 3415.

Butler, W. T. (1970a), *Biochemistry 9*, 44.

Butler, W. T. (1970b), *in* Chemistry and Molecular Biology of the Intercellular Matrix, Tristram, G. R., and Balazs, E. A., Ed., London, Academic Press, p 149.

Butler, W. T., Piez, K. A., and Bornstein, P. (1967), *Biochemistry 6*, 3771.

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

Fietzek, P. P., and Piez, K. A. (1969), *Biochemistry 8*, 2129.

Goldstein, A., and Adams, E. (1970), *J. Biol. Chem. 245*, 5478.

Grimm, L., and Grassmann, W. (1964), *Hoppe-Seylers' Z. Physiol. Chem. 337*, 161.

Gross, E. (1967), *Methods Enzymol. 11*, 238.

Hannig, K., and Nordwig, A. (1967), *in* Treatise on Collagen, Vol. 1, Ramachandran, G. N., Ed., London, Academic Press, p 73.

Highberger, J. H., Kang, A. H., and Gross, J. (1970), *Fed. Proc., Fed. Amer. Soc. Exp. Biol. 29*, 534.

Hill, R. L. (1965), *Advan. Protein Chem. 20*, 37.

Hodge, A. J. (1967), *in* Treatise on Collagen, Vol. 1, Ramachandran, G. N., Ed., London, Academic Press, p 185.

Kang, A. H., Bornstein, P., and Piez, K. A. (1967), *Biochemistry 6*, 788.

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

Kang, A. H., Piez, K. A., and Gross, J. (1969), *Biochemistry 8*, 1506.

Light, A. L. (1967), *Methods Enzymol. 11*, 417.

Michaeli, D., Martin, G. R., Kettman, J., Benjamini, E., Leung, D. Y. K., and Blatt, B. A. (1969), *Science 166*, 1522.

Miller, E. J., Lane, J. M., and Piez, K. A. (1969), *Biochemistry 8*, 30.

Miller, E. J., and Matukas, V. J. (1970), *in* Chemistry and Molecular Biology of the Intercellular Matrix, Tristram, G. R., and Balazs, E. A., Ed., London, Academic Press, p 109.

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

Prockop, D. J. (1970), *in* Chemistry and Molecular Biology of the Intercellular Matrix, Tristram, G. R., and Balazs, E. A., Ed., London, Academic Press, p 335.

Schroeder, W. A. (1967), *Methods Enzymol. 11*, 351.

Timpl, R., Fietzek, P. P., Furthmayr, H., Meigel, W., and Kühn, K. (1970), *FEBS (Fed. Eur. Biol. Soc.) Lett. 9*, 11.