Comparative Sequence Studies of Rat Skin and Tendon Collagen. I. Evidence for Incomplete Hydroxylation of Individual Prolyl Residues in the Normal Proteins*

Paul Bornstein†

ABSTRACT: Two homologous 36 amino acid containing peptides were obtained by CNBr cleavage of the $\alpha 1$ chains of rat skin and tail tendon collagen. The amino acid compositions of the two peptides differed only in that the skin peptide contained five hydroxyprolyl and seven prolyl residues per molecule whereas the tendon peptide contained three hydroxyprolyl and nine prolyl residues per molecule. The primary structures of the two peptides were found to be identical, apart from differences relating to the proline to hydroxyproline conversion. Both peptides consisted of triplets in which glycine always occupied position 1. Six of the twelve imino acid residues were in position 3 and these alone were susceptible to hydroxylation. However, hydroxylation was incomplete in both peptides, the degree of hydroxylation of several residues being greater in the peptide from skin collagen. The data indicate that the surrounding amino acid sequence plays a role in directing the hydroxylation of proline, but that tissue-specific factors are important in controlling the extent to which this hydroxylation takes place. Age-correlated changes in the degree of hydroxylation of proline were not observed. Whether the heterogeneity in collagen which results from the incomplete hydroxylation of proline is limited to the region of the α chains from which these peptides were derived (near the NH_2 terminus) or extends throughout the chains has not been established.

As a consequence of these studies information was also obtained regarding the specificity of bacterial collagenase. The minimal substrate requirement for collagenase appeared to be the sequences Gly-X-Y-Gly-Pro-Z or Gly-X-Y-Gly-Z-Hyp. (In both instances the Y–Gly bond was cleaved.) The findings point to the triplet on the carboxy side of the potentially susceptible bond in the collagen chain as the sequence primarily responsible for the specificity of bacterial collagenase.

The presence of hydroxyproline in collagen poses an unusual problem in the understanding of the synthesis of this protein. Although evidence exists for the presence of this imino acid in tRNA fractions, hydroxyprolyl-tRNA does not appear to be a precursor of collagen hydroxyproline (Urivetzky *et al.*, 1966). The available data strongly favor the view that prolyl residues in collagen are hydroxylated after their incorporation in peptide linkage (see Udenfriend (1966) for review). In the absence of a triplet code for hydroxyproline, a means of designating susceptible prolyl residues must exist if the locus of hydroxylation is to be specific.

A number of experiments have suggested that the surrounding amino acid sequence is at least partly responsible for directing the hydroxylation of proline in collagen. Early structural studies, based on partial acid hydrolysates, indicated that the sequences Gly-Pro, Pro-Hyp, and Hyp-Gly were common in collagen, whereas Gly-Hyp sequences were not found and Pro-Gly sequences were uncommon (Schroeder *et al.*, 1954; Kroner *et al.*, 1955). Greenberg *et al.* (1964) performed sequential Edman degradation on mixtures of collagenase peptides from carp swim bladder collagen. Hydroxyproline was found only in position 3 in the collagen triplet (in which glycine occupies position 1). Of the total proline, 82% could be accounted for in position 2 and only 1 prolyl residue/1000 was found in position 3. Finally, several groups of investigators (Schrohenloher *et al.*, 1959; Nagai, 1961; Gallop and Seifter, 1962; Grassman *et al.*, 1963) have recovered a substantial part of the collagen molecule, after collagenase digestion, as the peptides Gly-Pro-Hyp and Gly-Pro-Ala. These experiments therefore indicated that when proline occurred in position 3 in the collagen triplet the residue was hydroxylated. Rao and Harrington (1966), employing theoretical considerations, have performed calculations which are consistent with these data. It was shown that for vertebrate collagens the observed denaturation temperature (T_m) fitted that calculated for the triple helical structure proposed by Ramachandran and associates (1963) best if proline

* From the Laboratory of Biochemistry, National Institute of Dental Research, National Institutes of Health, Bethesda, Maryland, and the Research Training Unit, Department of Medicine, University of Washington School of Medicine, Seattle, Washington. *Received June 21, 1967.* Supported in part by Grant No. 11248 from the National Institute of Arthritis and Metabolic Diseases, U. S. Public Health Service. A portion of this work has appeared in preliminary form (Bornstein, 1967).

† Present address: Department of Medicine, University of Washington School of Medicine, Seattle, Wash.

and hydroxyproline were assigned to positions 2 and 3, respectively.

Recently an opportunity occurred to study in detail the structural basis for the hydroxylation of proline in collagen. During the course of experiments designed primarily to investigate the nature and location of cross-links in collagen, α1 chains of rat skin collagen were isolated by carboxymethylcellulose chromatography and cleaved with cyanogen bromide (CNBr). A peptide containing 36 amino acids, α1-CB2, was isolated by chromatography of the resulting peptide mixture on phosphocellulose (Bornstein *et al.*, 1966a; Bornstein and Piez, 1966). This peptide contained 12 glycyl, 7 prolyl, and 5 hydroxyprolyl residues per residue of methionine-derived homoserine and its molecular weight of 3230, determined by high-speed sedimentation equilibrium, agreed well with the minimum molecular weight of 3277 obtained from its amino acid composition. Additional experiments (Bornstein and Piez, 1966; Bornstein *et al.*, 1966b) indicated that α1-CB2 originated from a sequence near the NH_2 terminus of the α1 chain and comprised residues 16–51 of that chain. Similar experiments have now been performed with rat tail tendon collagen. In addition to the NH_2-terminal sequence, which differs from that in rat skin collagen (P. Bornstein, in preparation), a peptide homologous to that of α1-CB2 from rat skin collagen was isolated. This peptide had the same chromatographic properties on phosphocellulose and its amino acid composition was identical with that of α1-CB2 from rat skin except that it contained nine prolyl and three hydroxyprolyl residues per residue of homoserine.

In view of their amino acid compositions it seemed unlikely that the primary structures of the two peptides differed prior to hydroxylation. The above findings therefore suggested that factors in addition to the amino acid sequence were involved in the control of the hydroxylation of proline. In order to pursue this matter the amino acid sequences of the two peptides from normal rat skin and tendon collagen were determined. The supposition that the surrounding amino acid sequence is partly responsible for directing the hydroxylation of prolyl residues is supported by the data to be described. However, the finding of both position- and tissue-dependent incomplete hydroxylation of individual residues poses new problems in regard to an understanding of the manner in which the hydroxylation process is controlled.

Materials and Methods

Preparation of Collagen. Rat skin and tail tendon collagens were obtained from 6-week-old (100–500 g) male Sprague–Dawley rats, except as otherwise indicated. Both skin and tendon collagens were extracted and purified in the same fashion as previously described for skin collagen (Bornstein and Piez, 1964, 1966).

Preparation of α1 Chains and Isolation of Peptide α1-CB2.[1] Heat-denatured salt- and acid-extracted skin collagen and acid-extracted tendon collagen were fractionated on columns of carboxymethylcellulose (Piez *et al.*, 1963; Bornstein and Piez, 1966). Isolated α1 chains were usually contaminated with variable but small quantities of their covalently cross-linked dimer, β_{11}. Lyophilized α1 chains were dissolved at a concentration of 5 mg/ml in 0.1 M HCl. Cleavage with CNBr (Eastman Organic Chemicals) was performed under nitrogen at 30° for 4 hr employing a 100-fold M excess (relative to methionine) of CNBr. The reagent and solvent were removed by lyophilization and the resulting peptide mixture was chromatographed on phosphocellulose (Bornstein and Piez, 1966). Peptide α1-CB2 could be isolated in relatively pure form under the conditions used, the only contamination resulting from the presence of small quantities (5% or less) of peptide β_{11}-CB1. Phosphocellulose chromatograms of CNBr digests of α1 from skin and tendon were the same except for differences which were related to the NH_2-terminal peptide (α1-CB1). α1-CB2 was separated from buffer salts by chromatography on Bio-Gel P-2 (100–200 mesh, Bio-Rad Laboratories) equilibrated with 0.03 M ammonium propionate buffer (pH 4.5) (Bornstein and Piez, 1966).

Enzyme Hydrolyses. Digestion with trypsin and chymotrypsin was performed in 0.2 M NH_4HCO_3 (pH 7.8) containing 1×10^{-3} M $CaCl_2$. The enzymes were added as a 0.5% solution in 1×10^{-3} M HCl in a molar enzyme to substrate ratio of 1:200. Chymotryptic digestion was limited to 1 hr at room temperature while tryptic digestion was allowed to proceed for 3 hr. Digestions were terminated by lyophilization.

Papain (Worthington) was obtained as a crystalline suspension in 0.05 M sodium acetate (pH 4.5). The suspension was washed twice at 4° with 10% NaCl and dissolved in cold H_2O. Protein concentration was determined using $E^{1\%}_{1\,cm.\,280}$ 25. Digestion was performed in 0.07 M (in pyridine) pyridine acetate buffer, pH 5.5 containing 0.075 M BAL[2] at 40° for 40 hr. The concentration of enzyme was 2–5% that of the substrate. Digestion was terminated by lyophilization. A resinous material, resulting from polymerization of BAL, was frequently encountered but did not interfere with the subsequent separation of the products of papain digestion.

Crude collagenase (Worthington) was purified by the alumina C_γ gel method of Seifter *et al.* (1959). The purified preparation was inactive against casein, even after incubation at 37° for 24 hr, and contained an activity of 100 units/mg of protein (Seifter and Gallop, 1962). Digestion was carried out at 37° for 24 hr in 0.2 M NH_4HCO_3, containing 1×10^{-3} M $CaCl_2$ (pH 7.8) at an enzyme concentration of 0.05 mg/ml. The substrate concentration varied from 1 to 4 mg/ml. A drop of toluene was added to inhibit

[1] The peptides from both skin and tendon collagen will be referred to as α1-CB2.

[2] Abbreviations used: BAL, 2,3-dimercapto-1-propanol; dansyl or DNS, 1-dimethylaminonaphthalene-5-sulfonyl; PITC, phenyl isothiocyanate; TFA, trifluoroacetic acid; Hse, homoserine; PTH, phenylthiohydantoin derivative.

bacterial growth. The reaction was terminated by lyophilization.

A suspension of carboxypeptidase A (diisopropylphosphorofluoridate treated, Worthington) was dissolved at a concentration of 5 mg/ml in 2 M NH_4HCO_3 (pH 7.8). Digestion was carried out at 37° for 16 hr in 0.2 M NH_4HCO_3 (pH 7.8) at an enzyme concentration of 0.5 mg/ml. The reaction was terminated by lyophilization.

NH_2-Terminal Analysis and Sequential Degradation from the NH_2-Terminal End. NH_2-terminal amino acids in peptides were identified as their dansyl derivatives. Dansyl-Cl (B grade, California Corp. for Biochemical Research) was dissolved in isooctane and filtered. The filtrate was dried yielding orange crystals which were completely soluble in acetone. The procedure used was a modification of that described by W. R. Gray (manuscript in preparation).[3] A solution of peptide (5–50 mμmoles) was transferred to a 6 × 50 mm test tube and dried *in vacuo*. The peptide was dissolved in 15 μl of 0.1 M triethylamine carbonate buffer (pH 9.5) and an equal volume of 0.5% dansyl-Cl in acetone was added with thorough mixing. At the conclusion of the reaction the peptide was hydrolyzed in 6 N HCl under nitrogen at 108° for 18 hr. When proline or hydroxyproline was anticipated as NH_2-terminal hydrolysis was limited to 6–8 hr. Dansyl-amino acids were identified by thin layer chromatography on silica gel plates. Since all peptides examined were small and of known composition, a single chromatogram was sufficient to identify the dansyl-amino acid unequivocally. Benzene–pyridine–acetic acid (16:4:1) (Deyl and Rosmus, 1965) was used most frequently. Occasionally chromatography was performed in methyl acetate–isopropyl alcohol–concentrated NH_3 (9:7:4) (Seiler and Wiechmann, 1964). The latter system had the advantage of greater sensitivity but the resolution of the apolar dansyl-amino acids was limited.

Sequential degradation of a peptide using the Edman reagent together with the dansyl-Cl procedure was used to establish the sequence of a peptide. The method was essentially that of W. R. Gray (manuscript in preparation). In this method the new NH_2-terminal amino acid released after each step in the degradation is identified rather than the resulting PTH-amino acid. PITC, pyridine, and TFA were distilled prior to use. Coupling with PITC was performed at 37° for 3 hr by adding 100 μl of 5% PITC in pyridine (v/v) to the peptide dissolved in 200 μl of 50% aqueous pyridine. This and all subsequent procedures were performed under nitrogen in special tubes described by Taniuchi and Anfinsen (1966). After cyclization with TFA the nonvolatile residue was dissolved in a small amount of H_2O. An aliquot was removed for dansylation and the remainder was again dried and subjected to the second step in the degradation. At this point it was frequently necessary to adjust the pH of the 50% pyridine solution to 9–9.5 with a small amount of aqueous triethylamine prior to the addition of PITC.

Ion-Exchange Chromatography. Dowex 1-X2 (AG 1-X2, 200–400 mesh) and Dowex 50-X2 (AG 50W-X2, 200–400 mesh) were products of Bio-Rad Laboratories. Dowex 1-X2 was washed with 1 N HCl, water, 1 N NaOH, water, 2 N acetic acid, and finally with starting buffer (2 N pyridine, pH 8.0). The resin was packed in segments in a 55 × 0.9 cm column, jacketed at 37°. Elution was carried out with 100 ml of starting buffer followed by a linear gradient of 350 ml of 0.1 N acetic acid into 350 ml of starting buffer. A flow rate of 18 ml/hr was used and 3-ml fractions were collected. Dowex 50-X2 was washed with 2.5 N NaOH, water, 2 N HCl, water, and then with starting buffer (0.2 M (in pyridine) pyridine acetate, pH 3.1). Elution was carried out with 100 ml of starting buffer followed by a linear gradient of 350 ml of 2.23 M (in pyridine) pyridine acetate (pH 5.4) into 350 ml of starting buffer. Other conditions were the same as those described for Dowex 1-X2 chromatography except that the column was jacketed at 40°. Peptides were identified in effluent fractions by performing manual ninhydrin analysis of the unhydrolyzed specimens.

Phosphocellulose chromatography of enzymatic fragments of α1-CB2 was performed as described for CNBr digests of collagen chains (Bornstein and Piez, 1966).

Paper Chromatography and Electrophoresis. Descending paper chromatography was performed on Whatman No. 1 paper strips in butanol–acetic acid–water (12:3:5). Peptides were identified with the ninhydrin reagent. Preparative paper chromatography was performed by streaking the sample at the origin to within 0.25 in. of the edge of the paper. Peptides were identified by cutting and staining 0.5-in. guide strips from both sides of the paper and were eluted with 1% acetic acid at 50°.

Conventional paper electrophoresis was performed in 0.5 M acetic acid (pH 2.7) on Whatman No. 3MM paper at 14 v/cm for 3 hr. High-voltage paper electrophoresis was performed on 35 × 55 cm sheets of Whatman No. 3MM filter paper in pyridine–acetic acid–water (1:10:289) (pH 3.7) at 2000 v for 120 min and on 23.5 × 104 cm sheets of Whatman No. 3MM filter paper in 4% formic acid at 4000 v for 140 min. The cadmium–ninhydrin stain (Blackburn, 1965) was used for greater sensitivity. Preparative paper electrophoresis was performed as described for paper chromatography.

Automatic Amino Acid and Peptide Analysis. Amino acid analyses were performed on a single-column automatic instrument modified for high-speed analysis (Miller and Piez, 1966). Peptides were hydrolyzed under nitrogen in 6 N HCl at 108° for 24 hr. A correction of 4.9% was applied for loss of serine. With temperature programming and a slightly modified gradient homoserine eluted between serine and glutamic acid and was well separated from both amino acids. Homoserine lactone eluted at 173 min between histidine (154 min) and arginine (210 min). Peptides were ana-

[3] I wish to thank Dr. W. R. Gray for making manuscripts describing the dansyl-Cl and Edman procedures available to me prior to publication.

lyzed in the same manner as amino acids. The commercially prepared synthetic tripeptides Gly-Pro-Hyp and Gly-Pro-Pro (Cyclo Chemical Corp.) were kindly provided by Dr. S. Udenfriend. The amino acid composition of Gly-Pro-Hyp, in residues per peptide was $Gly_{1.04}$,$Pro_{0.98}$,$Hyp_{0.98}$, and that of Gly-Pro-Pro, $Gly_{0.98}$,$Pro_{2.10}$. Gly-Pro-Pro was contaminated with approximately 6% proline but the peptides were free of other ninhydrin-reacting substances and their sequences were verified by dansylation and Edman degradation. Gly-Pro-Hyp and Gly-Pro-Pro eluted at 83 and 100 min, respectively, in positions normally occupied by cystine and isoleucine. The integration constant calculated for Gly-Pro-Hyp was 21.0 and that for Gly-Pro-Pro was 17.4 (leucine, 25.3).

Experiments with Iron-Deficient Rats. Male Sprague–Dawley rats (4–5-weeks old, 70–80 g) were placed on an iron-deficient diet described by McCall *et al.* (1962) and prepared commercially by General Biochemicals. The diet contained 1–3 mg of iron/kg. Animals were housed in plastic cages and provided with glass-distilled water. After 6 weeks on the diet, the animals were sacrificed and skin collagen was prepared as described previously.

Results

The Amino Acid Sequence of α1-CB2. Separation of Enzymatic Fragments by Phosphocellulose and Sephadex Chromatography. The amino acid sequence of α1-CB2 was found to be identical in the peptides from skin and tendon collagen with the exception of differences resulting from the hydroxylation of proline. The proposed sequence for α1-CB2 is illustrated in Figure 1. Although positions 9, 12, 15, 24, 27, and 30 are designated as hydroxyproline, hydroxylation was incomplete in all of these positions. The variation in the degree of hydroxylation, both in regard to a particular position and to the tissue from which the peptide was obtained, will be described.

The first step in the degradation of α1-CB2 consisted of digestion with either trypsin, chymotrypsin, or collagenase. In the following experiments the same results were obtained with the peptide from skin and tendon collagen, except as indicated. The products of tryptic digestion were chromatographed on phosphocellulose. Two well-separated peaks were observed. The amino acid composition of the more basic peptide (α1-CB2-T1, Table I) indicated that it contained six amino acid residues including arginine. On the basis of the specificity of trypsin, arginine is therefore located in position 6 in α1-CB2. The analyses of α1-CB2-T1 and α1-CB2-T2 (Table I) indicated that proline but not hydroxyproline was present in the first tryptic fragment and that the difference in the hydroxyproline and proline contents of the peptides from the two tissues was limited to residues 7–36.

The products of chymotryptic digestion of α1-CB2 were also chromatographed on phosphocellulose (Figure 2). The first major peak (α1-CB2-C2) contained homoserine and this peptide therefore constituted the

TABLE I: The Amino Acid Composition of Peptide α1-CB2 and Its Enzymatic Fragments from Rat Skin and Tail Tendon Collagen.[a]

	Skin α1-CB2-						Tendon α1-CB2-				
	T1 (1–6)	T2 (7–36)	C1 (1–20)	C2 (21–36)	Col7 (22–33)		T1 (1–6)	T2 (7–36)	C1 (1–20)	C2 (21–36)	Col7 (22–33)
4-Hydroxyproline	4.9	—	5.1	2.5	2.5	2.3	3.0	—	2.9	2.3	0.5	0.5
Serine	2.2	1.0	1.0	1.1	1.0	1.0	2.0	1.0	1.0	1.0	0.9	1.0
Homoserine[b]	1.0	—	1.0	—	1.0	—	0.9	—	1.0	—	1.0	—
Glutamic acid	4.0	—	4.0	1.1	3.0	2.1	4.0	—	4.0	1.1	3.2	2.0
Proline	6.9	2.0	5.0	4.3	2.4	1.3	8.5	2.0	6.7	4.3	4.2	3.2
Glycine	12.1	1.9	10.2	6.8	5.0	3.9	11.8	2.1	10.0	7.0	5.0	3.9
Alanine	2.1	—	2.0	1.0	1.0	1.0	2.1	—	2.1	1.0	1.0	1.0
Leucine	1.0	—	1.0	1.0	—	—	1.0	—	1.0	1.0	—	—
Phenylalanine	0.9	—	1.0	0.9	—	—	1.0	—	1.0	0.9	—	—
Arginine	1.0	1.0	—	1.0	—	—	1.0	1.0	—	1.0	—	—

[a] The values are given as residues per peptide and are the averages of two or more determinations. A dash indicates that the amino acid was either entirely absent or present as less than 0.1 residue. T, C, and Col are tryptic, chymotryptic, and collagenase fragments, respectively. The numbers in parentheses refer to the residue numbers in Figure 1.
[b] Includes homoserine lactone.

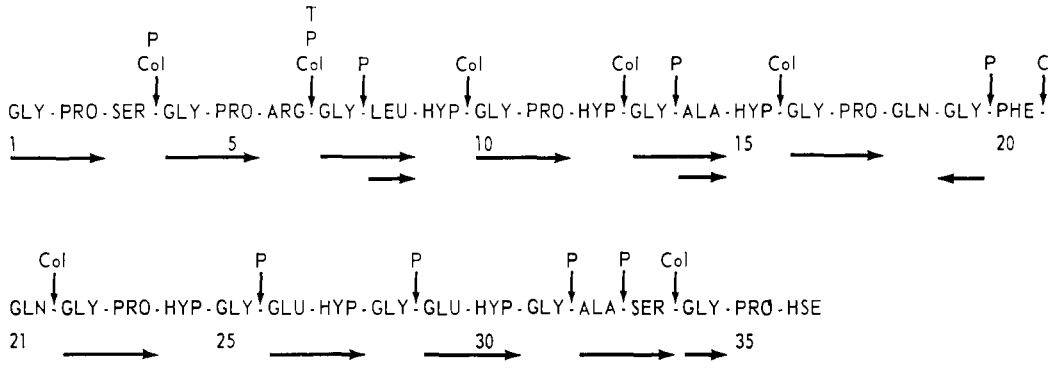

FIGURE 1: Proposed amino acid sequence of α1-CB2 and positions of cleavage by: Col, collagenase; P, papain; T, trypsin; and, C, chymotrypsin. Only the primary sites of papain cleavage of the individual chymotryptic peptides are indicated. Positions 9, 12, 15, 24, 27, and 30 are designated as hydroxyproline but were incompletely hydroxylated in both skin and tendon α1-CB2, the degree of hydroxylation of several positions being greater in the peptide from skin collagen (see text). The horizontal arrows (→) indicate the extent of Edman degradations. In one instance carboxypeptidase A (←) was used to establish the presence of glycine at position 19.

COOH-terminal portion of α1-CB2. The observed chromatographic heterogeneity of α1-CB2-C2 was due to the presence of a small amount of tryptic activity in the chymotrypsin preparation, resulting in cleavage of α1-CB2-C1 at the arginyl residue. One of the two peptides so produced, α1-CB2-C1-T2 (residues 7–20), chromatographed just before α1-CB2-C2. Possibly the equilibrium between homoserine and homoserine lactone and the consequent charge difference between the two forms of α1-CB2-C2 contributed to the chromatographic heterogeneity of the peptide. The amino acid compositions of α1-CB2-C1 and α1-CB2-C2 (Table I) together with the specificity of chymotrypsin place phenylalanine in position 20 in α1-CB2. The predominant difference in the hydroxyproline and proline contents of α1-CB2 from skin and tendon was present in α1-CB2-C2, residues 21–36. However, a significant difference also existed in α1-CB2-C1 (see below).

Collagenase digestion was performed either on the whole peptide (α1-CB2) or on the chymotryptic fragment α1-CB2-C1 (residues 1–20) previously separated by phosphocellulose chromatography. In both instances the products were chromatographed on Sepha-

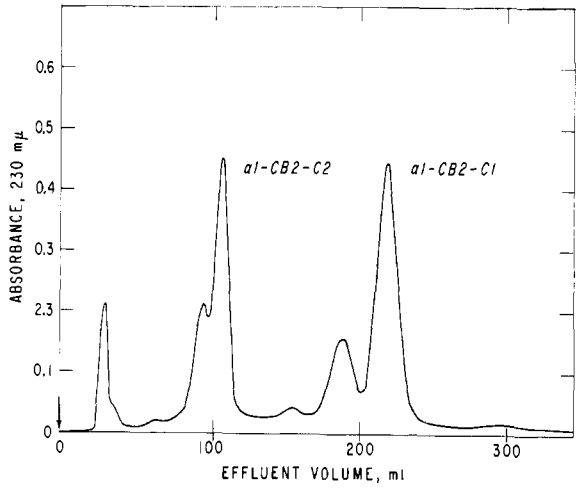

FIGURE 2: Phosphocellulose gradient elution pattern of peptides obtained by chymotryptic cleavage of α1-CB2. Chromatography was performed at pH 3.8, 40°. The arrow indicates placement of the sample dissolved in 5 ml of starting buffer, 0.001 M sodium acetate (pH 3.8). Fractions of 10 ml were collected.

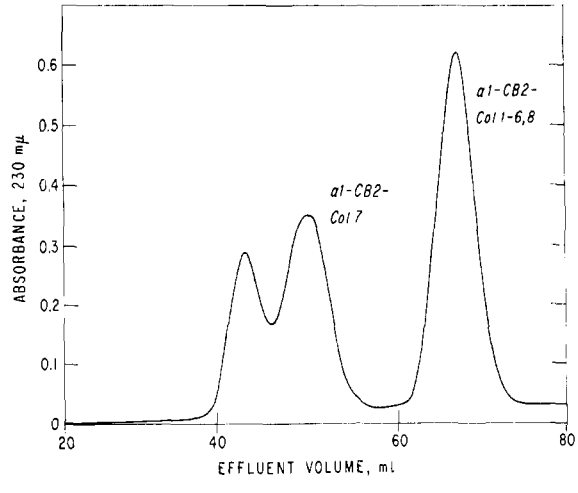

FIGURE 3: Sephadex G-25 elution pattern of peptides obtained by collagenase digestion of α1-CB2. Chromatography was performed in 0.03 M ammonium propionate (pH 4.5). Fractions of 3 ml were collected.

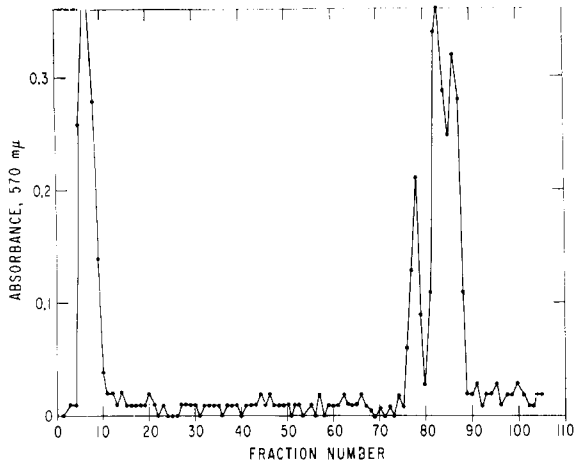

FIGURE 4: Dowex 1-X2 chromatogram of the small peptides separated by Sephadex G-25 (60–75-ml effluent volume, Figure 3) after collagenase digestion of α1-CB2. Conditions of chromatography are reported in the text. Fractions of 3 ml were collected and ninhydrin analysis of each fraction was performed.

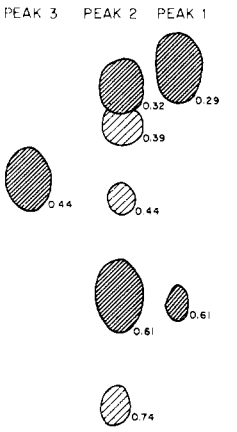

FIGURE 5: Descending paper chromatogram (butanol–acetic acid–water, 12:3:5) of the three peaks obtained by Dowex 50-X2 chromatography of a collagenase digest of α1-CB2. The peptides had previously been chromatographed on Sephadex G-25 and Dowex 1-X2 and only fractions 80–89 of the latter chromatogram (Figure 4) were used in the Dowex 50-X2 chromatogram. Lightly shaded spots represent minor components. R_F values are indicated.

dex G-25. In the case of α1-CB2 (Figure 3) the first peak, eluting at the void volume of the column, contained the collagenase together with a small amount of peptide β_{11}-CB1 which contaminated α1-CB2 and was not cleaved by the enzyme. The predominant peptide in the second peak, α1-CB2-Col7, had an amino acid composition similar to that of α1-CB2-C2 (Table I) and therefore originated from the COOH-terminal half of α1-CB2. The third peak, which was totally retarded, contained a mixture of small peptides representing the remainder of the sequence of α1-CB2. The chromatogram of a collagenase digest of α1-CB2-C1 (residues 1–20) differed from that of the digest of the whole peptide in that, as expected, the peak containing α1-CB2-Col7 was missing.

The small peptides derived from α1-CB2, α1-CB2-C1, and α1-CB2-C2 by collagenase and papain digestion were further separated by a combination of ion-exchange chromatography, paper chromatography, and paper electrophoresis. The amino acid sequences of these peptides and their relative positions in the parent molecule are described in detail below. In most instances three criteria of purity were employed: amino acid composition, chromatographic homogeneity on paper, and chromatographic homogeneity of the unhydrolyzed peptide on the automatic amino acid analyzer.

The Amino Acid Sequence of α1-CB2-C1. RESIDUES 1–3. GLY-PRO-SER. A collagenase digest of α1-CB2 was chromatographed on Sephadex G-25 (Figure 3) and the small peptides in this digest were then separated on Dowex 1-X2 (Figure 4). Fractions 80–89 of the latter elution pattern were chromatographed on Dowex 50-X2 and resolved into three peaks. Paper chromatography of the second peak revealed five ninhydrin-positive spots (Figure 5). The amino acid analysis[4] of the peptide with an R_F of 0.32 was $Gly_{1.10}Pro_{0.93}Ser_{0.98}(Glu_{0.11})$; Edman degradation: (step 1) DNS-glycine and (step 2) DNS-proline. Since the only seryl residue in α1-CB2-C1 was located in the tryptic peptide α1-CB2-T1 (residues 1–6), Gly-Pro-Ser must constitute the NH_2-terminal triplet in α1-CB2.

RESIDUES 4–6. GLY-PRO-ARG. This tripeptide contained the only basic residue in α1-CB2 and was isolated essentially pure by chromatography of the small collagenase-derived peptides (previously separated by Sephadex G-25) on either Dowex 1-X2 (Figure 4) or Dowex 50-X2 (Figure 6). In the former case the peptide, together with NH_4^+, eluted at the start of the chromatogram. On Dowex 50-X2 the peptide eluted in fractions 86–90; amino acid analysis: $Gly_{1.05}Pro_{0.96}Arg_{0.99}$; Edman degradation: (step 1) DNS-glycine and (step 2) DNS-proline. The position of this tripeptide was consistent with the results of tryptic cleavage of α1-CB2 (Table I and Figure 1).

RESIDUES 7–9. GLY-LEU-HYP*.[5] As in the case of Gly-Pro-Ser, the major form of this tripeptide (Gly-Leu-Hyp) was isolated by Dowex 1-X2 and Dowex 50-X2 chromatography followed by paper chromatography in butanol–acetic acid–water. Its R_F in the latter system was 0.61 (Figure 5); amino acid analysis: $Gly_{0.99}Leu_{1.00}Hyp_{0.99}$; Edman degradation: (step 1) DNS-glycine and (step 2) DNS-leucine. The location

[4] Amino acid analyses are reported as residues per peptide. Residues thought to be impurities are in parentheses. Amino acids accounting for less than 0.1 residue are omitted.

[5] Hyp* is used here and subsequently to indicate the presence of incompletely hydroxylated prolyl residues.

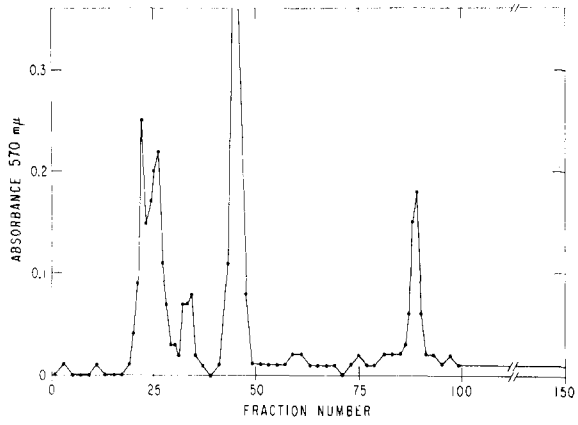

FIGURE 6: Dowex 50-X2 chromatogram of the small peptides separated by Sephadex G-25 (60–75-ml effluent volume, Figure 3) after collagenase digestion of α1-CB2. Conditions of chromatography are reported in the text. Fractions of 3 ml were collected and ninhydrin analysis of each fraction was performed.

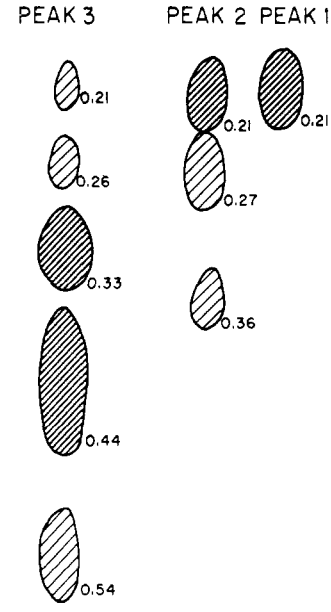

FIGURE 7: Descending paper chromatography (butanol–acetic acid–water, 12:3:5) of the three peaks obtained by Dowex 1-X2 chromatography of a papain digest of α1-CB2-C1. Lightly shaded spots represent minor components. R_F values are indicated.

of Gly-Leu-Hyp* in positions 7–9 was demonstrated as follows. α1-CB2-C1 was cleaved with papain and the digest chromatographed on Dowex 1-X2 and subsequently on paper (Figure 7). The peptide mixture was complex since papain characteristically did not cleave any bond completely and additional heterogeneity results from incomplete hydroxylation of prolyl residues in positions 9, 12, and 15 (see below). However, the amino acid compositions of two peptides served to locate the tripeptide Gly-Leu-Hyp* as well as the succeeding two collagenase-derived tripeptides, Gly-Pro-Hyp* and Gly-Ala-Hyp* in the sequence of α1-CB2. The amino acid composition of the peptide in peak 3 (Dowex 1-X2) with an R_F of 0.44 (Figure 7) was $Leu_{1.00}Hyp_{1.88}Gly_{2.14}Pro_{0.98}$. DNS-leucine was found by dansylation. The amino acid composition of the peptide in peak 2 (Dowex 1-X2) with an R_F of 0.21 (Figure 7) was $Ala_{0.99}Hyp_{1.04}Gly_{1.96}Pro_{0.97}Glu_{1.02}$. DNS-alanine was found by dansylation. Since the glutaminyl residue in α1-CB2-C1 was located among the last five residues (see below), the relative positions of the two papain fragments were fixed and the partial sequence of this region could be written as Leu-Hyp*-(Hyp,-Pro,Gly_2)-Ala-(Hyp,Glu,Pro,Gly_2).

RESIDUES 10–12. GLY-PRO-HYP*. Both forms of this tripeptide (Gly-Pro-Hyp and Gly-Pro-Pro) chromatographed as a single peak on Dowex 1-X2 (fractions 75–79, Figure 4) well separated from the other neutral collagenase-derived peptides. The amino acid analysis of this peak differed significantly in the case of the peptide from skin and tendon collagen (see below). The composition of the peptide mixture from skin collagen was $Gly_{1.03}Pro_{1.09}Hyp_{0.74}$; Edman degradation: (step 1) DNS-glycine and (step 2) DNS-proline. These tripeptides were assigned to positions 10–12 (Figure 1) as a result of the analyses of the papain-produced fragments of α1-CB2-C1.

RESIDUES 13–15. GLY-ALA-HYP*. The major form of this collagenase-produced peptide (Gly-Ala-Hyp) was also eluted in fractions 80–89 by Dowex 1-X2 chromatography (Figure 4) and was localized to the first peak in the subsequent Dowex 50-X2 fractionation of these peptides. Paper chromatography of this peak (Figure 5) indicated that it was almost homogeneous (R_F 0.29) with the exception of a small amount of material migrating in the position of Gly-Leu-Hyp (R_F 0.61); amino acid analysis: $Gly_{1.03}Ala_{1.00}Hyp_{0.97}$; Edman degradation: (step 1) DNS-glycine and (step 2) DNS-alanine. This tripeptide was assigned to positions 13–15 (Figure 1) as a result of the analyses of the papain-produced fragments of α1-CB2-C1.

RESIDUES 16–21. GLY-PRO-GLN-GLY-PHE-GLN. This peptide was isolated by Dowex 50-X2 chromatography (fractions 32–36, Figure 6) of a collagenase digest of α1-CB2. Alternatively when α1-CB2-C1 (residues 1–20) was the starting material, a peptide with similar chromatographic properties but containing one less glutaminyl residue was isolated. In both instances the hexa- or pentapeptide eluted with other neutral peptides on both Dowex 50-X2 and Dowex 1-X2 indicating that residues 18 and 21 were glutaminyl residues; amino acid analysis: $Gly_{1.96}Pro_{1.04}Glu_{1.95}$-$Phe_{0.99}$; Edman degradation: (step 1) DNS-glycine and (step 2) DNS-proline; residue analysis after step 2: $Gly_{1.39}Pro_{0.24}Glu_{2.21}Phe_{0.98}$. Phenylalanine and glutamine were placed in positions 20 and 21 by the points of cleavage of papain, chymotrypsin, and collagenase (Figure 1). The residue in position 19 was identified by carboxypeptidase A digestion of the

papain-produced fragment (residues 14–19). Incubation of 0.12 μmole of peptide with the enzyme yielded, after correction for the enzyme blank, 0.055 μmole of glycine. Glutamine in position 18 was not released, possibly because the penultimate prolyl residue in position 17 inhibited further degradation by carboxypeptidase A.

The Amino Acid Sequence of α1-CB2-C2. RESIDUES 21–36. The approach to the amino acid sequence of this portion of the molecule was not as systematic as that used for residues 1–20 largely because only a single bond, the Ser–Gly bond (residues 33–34), was susceptible to collagenase. The resulting tripeptide (residues 34–36) was isolated from collagenase digests of α1-CB2. Separation was achieved by sequential Dowex 1-X2, Dowex 50-X2, and paper chromatography. The amino acid composition of the peptide (peak 2, R_F 0.39, Figure 5) was $Gly_{1.12}Pro_{1.03}Hse_{0.86}$-$(Ser_{0.12})$; Edman degradation: (step 1) DNS-glycine. Since α1-CB2 resulted from CNBr cleavage at a methionyl bond, homoserine must be COOH terminal in position 36 and proline is therefore restricted to position 35.

The remainder of the sequence of α1-CB2-C2 was determined by the analysis of its papain fragments separated by Dowex 1-X2 chromatography and high-voltage paper electrophoresis at pH 1.9. The use of papain was helpful since in addition to providing overlapping sequences this enzyme characteristically did not cleave prolyl (or hydroxyprolyl) bonds or bonds formed by the imino group of these imino acids.

Dowex 1-X2 chromatography of a papain digest of residues 22–36 (obtained from a partial collagenase digest of α1-CB2) revealed three well-separated peaks eluting in the neutral region of the gradient. When the peptide was obtained from skin collagen, amino acid analysis of the first peak was $Gly_{2.00}Pro_{1.27}Hyp_{0.59}$. This peak could be shown to consist of two peptides by paper chromatography in butanol–acetic acid–water; R_F 0.29, $Gly_{1.90}Pro_{1.10}Hyp_{0.94}$; and R_F 0.36, $Gly_{2.00}Pro_{1.92}$. Edman degradation and dansylation of the mixture of peptides in this peak revealed: (step 1) DNS-glycine and (step 2) DNS-proline. No hydroxyproline was found in the second position. The amino acid composition of the second peak was $Ala_{1.05}Ser_{1.05}Gly_{1.10}Pro_{0.95}Hse_{0.86}$; Edman degradation: (step 1) DNS-alanine and (step 2) DNS-serine. The third peak contained a peptide with an amino acid composition consistent with the sequence Ser-Gly-Pro-Hse. The Dowex column was then eluted with 3 N acetic acid. Acidic peptides containing glutamic acid, glycine, proline, hydroxyproline, and alanine were eluted. These peptides were resolved by paper electrophoresis at pH 1.9 (see below).

When papain digestion was performed on α1-CB2-C2, similar peptides were isolated but the peptides containing glycine, proline, and hydroxyproline were found to contain, in addition, glutamine. On the basis of these data the sequence of residues 21–25 was presumed to be Gln-Gly-Pro-Hyp*-Gly and that of residues 32–36, Ala-Ser-Gly-Pro-Hse.

Paper electrophoresis, at pH 1.9, of the acidic peptides resulting from a papain digest of tendon α1-CB2-C2 revealed three ninhydrin-positive spots. Amino acid analysis of the peptide material from the most rapidly migrating and most prominent spot, peptide A, was $Glu_{0.99}Gly_{1.01}Pro_{0.83}Hyp_{0.10}$. Amino acid analyses of the other two peptides were: peptide B, $Glu_{0.99}Gly_{1.05}$-$Pro_{0.71}Hyp_{0.17}Ala_{0.97}$; and peptide C, $Glu_{0.98}Gly_{1.02}$-$Pro_{0.58}Hyp_{0.42}$. All three peptides contained NH_2-terminal glutamic acid by dansylation. Edman degradation of peptide A revealed proline and a small amount of hydroxyproline in the second position. Residue analyses after the first step were: peptide A, $Glu_{0.24}$-$Gly_{1.00}Pro_{0.87}Hyp_{0.18}$; peptide C, $Glu_{0.59}Gly_{1.00}Pro_{0.46}$-$Hyp_{0.46}$.

These data suggested the sequence Glu-Hyp*-Gly-Glu-Hyp*-Gly-Ala for positions 26–32 in α1-CB2. Peptide C appeared to be a dimer of the sequence Glu-Hyp*-Gly arising from incomplete cleavage by papain of the second Gly–Glu bond. Although the mixture of peptides contained only 20% of the total imino acids as hydroxyproline, the percentage of hydroxyproline in peptide C was consistently 40–50% and that in peptide A 10–20%, suggesting that cleavage of the second Gly–Glu bond was inhibited by the presence of hydroxyproline in either positions 27 or 30 (or both)

Since peptide A was degraded by the Edman reagent the presence of a γ-glutamyl bond in this peptide was unlikely. However, in view of the evidence for γ-glutamyl bonds in collagen (Franzblau *et al.*, 1963), this matter was further investigated by comparing the electrophoretic properties of the native peptide with that of synthetic α- and γ-glutamyl peptides. α- and γ-L-glutamyl-L-prolyl-glycine were synthesized by conventional techniques[6] and purified from contaminating precursor products by high-voltage electrophoresis at pH 3.7. The purified peptides were homogeneous by amino acid analyses and the sequence of the α-glutamyl peptide was verified by dansylation and Edman degradation. As expected the γ-glutamyl peptide contained NH_2-terminal glutamic acid but could not be degraded by PITC. The electrophoretic migration of the two peptides at pH 1.9, 2.7, and 3.7 differed in accord with the difference in pK of the α- and γ-carboxyl groups of glutamic acid. The electrophoretic migration of the native peptide Glu-Hyp*-Gly was identical with that of the synthetic α-glutamyl peptide at pH 1.9, 2.7, and 3.7 indicating that the Glu–Hyp* bonds at positions 26–27 and 29–30 were α-glutamyl linkages.

Evidence for Incomplete Hydroxylation of Proline in α1-CB2 from Both Skin and Tendon. INCOMPLETE HYDROXYLATION OF PROLINE AT POSITION 12. The only position for which direct quantitative data could be obtained regarding the extent of hydroxylation was position 12. Collagenase apparently cleaved the bonds

[6] I am greatly indebted to Dr. J. E. Folk, National Institute of Dental Research, for performing these syntheses.

between residues 9–10 and 12–13 completely. The resulting tripeptides (Gly-Pro-Hyp*, residues 10–12) were isolated as a mixture by Sephadex G-25 chromatography (Figure 3) followed by Dowex 1-X2 chromatography (Figure 4). The heterogeneity of the peak containing these tripeptides (fractions 75–79, Figure 4) was demonstrated by paper chromatography in butanol–acetic acid–water. Two spots with R_F values of 0.27 and 0.44 were detected by the ninhydrin reaction and yielded peptides with the compositions $Gly_{0.98}Pro_{1.11}Hyp_{0.90}$ and $Gly_{0.93}Pro_{2.13}$, respectively. Since quantitative recovery after elution from paper was difficult, the mixture of tripeptides obtained from the Dowex column was placed on the amino acid analyzer column and eluted by gradient elution. Two well-separated peaks were observed eluting at the same positions as the synthetic tripeptides (Gly-Pro-Hyp and Gly-Pro-Pro). The relative quantities of these tripeptides in a mixture from α1-CB2 could be determined using previously calculated integration constants. The results are summarized in Table II. The data indicate that proline in position 12 was hydroxylated to the extent of 70–80% in skin collagen and 50–60% tendon collagen. Since the same findings were obtained from α1-CB2 and α1-CB2-C1 the unlikely possibility that the two sequences Gly-Pro-Hyp and Gly-Pro-Pro arose from different positions in the α1-CB2 molecule can be excluded. In the case of skin collagen no significant differences were observed in the degree of hydroxylation of prolyl residue 12 in rats varying in age from 17 days to 1 year.

TABLE II: The Degree of Hydroxylation of Proline in Position 12 in α1-CB2 from Rat Skin and Tendon Collagen.

Source	Age of Rats (days)	% Gly-Pro-Hyp	% Gly-Pro-Pro
Skin, α1-CB2	17	72	28
Skin, α1-CB2[a]	40	81	19
Skin, α1-CB2	120	73	27
Skin, α1-CB2	360	73	27
Skin, α1-CB2-C1	40	78	22
Tendon, α1-CB2[b]	40	58	42
Tendon, α1-CB2-C1	40	53	47

[a] Average of four analyses. [b] Average of two analyses.

INCOMPLETE HYDROXYLATION OF PROLINE AT POSITIONS 9 AND 15. Collagenase digests of both skin and tendon α1-CB2 consistently contained several minor components. Analyses of the ninhydrin-positive spots with R_F values of 0.44 and 0.74 (peak 2, Figure 5) were $Gly_{1.06}Ala_{0.88}Pro_{1.08}$ and $Gly_{1.00}Leu_{1.07}Pro_{0.93}$, respectively. Since leucine is unique in α1-CB2 and only one of the two alanyl residues is adjacent to an imino acid (Figure 1) the above peptides must have originated from positions 7–9 and 13–15. Direct quantitation of the degree of hydroxylation of proline in positions 9 and 15 was not possible. However, in view of the hydroxyproline contents of α1-CB2-C1 from skin and tendon (Table I) and the known degree of hydroxylation at position 12 (Table II) hydroxylation at positions 9 and 15 must be 85–90% complete in the collagens from both tissues.

INCOMPLETE HYDROXYLATION OF PROLINE AT POSITIONS 24, 27, AND 30. The evidence for incomplete hydroxylation at these positions resulted from the finding that the amino acid compositions of the relevant peptides in papain digests of α1-CB2-C2 contained nonwhole number ratios of proline and hydroxyproline. (In the case of the peptides derived from positions 22–25 the two forms were also separated by paper chromatography.) Amino acid analyses of these mixtures indicated that in skin collagen proline in position 24 was hydroxylated to the extent of 50–60% and proline in positions 27 and 30 to the extent of 90% or more. In tendon collagen the degree of hydroxylation of proline in position 24 was 20–40% and that in positions 27 and 30, 10–20%. These figures are estimates. However, the hydroxyproline contents of α1-CB2-C2 from skin and tendon, 2.5 and 0.5 residues, respectively (Table I), are in good agreement with these estimates.

The Effect of Iron Deficiency on the Hydroxylation of Proline in α1-CB2. Peptide α1-CB2 was isolated from skin collagen of rats made iron deficient by restriction of dietary iron during growth. Prior to sacrifice the hematocrits of these rats averaged 31%, compared with values of 48–50% for control animals. α1-CB2 was cleaved with chymotrypsin and the resulting two peptides were isolated by phosphocellulose chromatography. The hydroxyproline content of α1-CB2-C1 was 2.4 and the proline content was 4.4 residues/peptide. The hydroxyproline and proline contents of α1-CB2-C2 were both 2.4 residues/peptide. These amino acid analyses are identical, within experimental error, with those of α1-CB2-C1 and α1-CB2-C2 from normal rat skin collagen (Table I). It is therefore apparent that the level of iron deficiency achieved during these experiments did not measurably influence the hydroxylation of proline in skin collagen in the region of the α1 chain represented by α1-CB2.

Discussion

The elucidation of the primary structures of two homologous 36 amino acid containing peptides from rat skin and tendon collagen has provided an opportunity to examine directly a number of aspects of collagen structure which could heretofore only be inferred from analyses of smaller enzymatic fragments or from studies of the whole molecule and its component chains. The primary question to which this study addressed itself was the explanation for the observation that the two peptides contained identical nonimino acid compositions and the same total imino

acids but different proline to hydroxyproline ratios. More fundamentally, information was sought regarding the structural basis for the control of the hydroxylation of proline in collagen.

The results which have been described indicate that the amino acid sequences of the two peptides are identical, apart from differences relating to the proline to hydroxyproline conversion. In addition, the prolyl residues which are candidates for hydroxylation are the same. However, hydroxylation of all prolyl residues was incomplete. The difference in hydroxyproline contents of the two peptides is explained by the finding that the degree of hydroxylation of several homologous prolyl residues differed in the two tissues. Although other tissue factors may be involved, this finding is probably due to differences in the activity of the enzyme system responsible for the hydroxylation of proline. Conceivably several proline hydroxylases are present in both skin and tendon. These hydroxylases could have different activities, and possess specificities which are dependent on the adjacent amino acid sequence of the collagen substrate. However, since both proline and lysine may be hydroxylated by the same enzyme (Kivirikko and Prockop, 1967b) it seems less likely that multiple enzymes are involved in the hydroxylation of proline.

The data presented here are partly in accord with previous experiments in that proline in position 2 in the collagen triplet was not hydroxylated in either skin or tendon collagen. Proline in position 3 was always susceptible to hydroxylation but was incompletely and variably hydroxylated. It would therefore appear that while the adjacent amino acid sequence directs the hydroxylation of proline, tissue-specific factors are important in controlling the extent to which this hydroxylation takes place. Since differences in the degree of hydroxylation of individual residues were found in the same peptide, the surrounding amino acid sequence may also influence the extent of hydroxylation, perhaps as a result of a different affinity of the enzyme for different sequences in collagen.

A consequence of the incomplete hydroxylation of proline in collagen is the heterogeneity of the final primary structure of the protein. If this incomplete hydroxylation were present throughout all three α chains the number of different sequences for the collagen molecule would be extremely large. Since amino acid analyses of the entire α chains of rat tendon and skin collagen do not differ significantly in hydroxyproline (Piez *et al.*, 1963; P. Bornstein, unpublished observations) the differences noted in α1-CB2 must either be limited to one region of the chain or else be compensated for by changes in the opposite direction elsewhere in α1.

In vitro experiments suggest that the polypeptide substrate for hydroxylation may be relatively large (Gottlieb *et al.*, 1966; Lukens, 1966; Juva and Prockop, 1966; Kivirikko and Prockop, 1967a). These experiments were performed under conditions in which normal hydroxylation of proline was inhibited and evidence that hydroxylation occurs on the growing polypeptide chain also exists (Manning and Meister, 1966). If, however, hydroxylation normally did not occur until the precursor collagen chains were largely complete, the hydroxylation process might be influenced by the conformation of the substrate. It has been shown previously that α1-CB2 is located within the first 5% of the α1 chain (Bornstein and Piez, 1966; Bornstein *et al.*, 1966b) and that the amino acid sequence on the amino side of α1-CB2 does not conform with that required for the formation of the collagen helix (Kang *et al.*, 1967). Incomplete hydroxylation of proline may therefore be limited to regions near the NH_2 termini of the α chains. Sequence analysis of peptides from other regions should resolve this question.

Since tissues from young growing rats were used for the majority of these studies, the possibility that the observed incomplete hydroxylation of proline was a consequenee of the immaturity of the animal was considered. Chvapil and Kobrle (1961) have reported data which indicate that the hydroxyproline content of rat lung and tendon collagen increases with age although the proline contents of these tissues were found to remain constant. These findings are puzzling since the data would require that the basic sequence of collagen changes with age. Other investigators (Pine and Holland, 1966) have not found such age-correlated changes in collagen composition. The data in Table II indicate that, at least in regard to proline in position 12 of skin α1-CB2, no significant difference in the degree of hydroxylation was detected over an age range from 17 days to 1 year.

The requirement for Fe^{2+} by the enzyme system which hydroxylates proline has been demonstrated (Hurych and Chvapil, 1965; Hutton *et al.*, 1967; Kivirikko and Prockop, 1967b). The possibility that rats on a standard laboratory diet might become iron deficient seemed too remote to consider. However, iron deficiency was induced in order to investigate whether the hydroxylation of prolyl residues which were normally incompletely hydroxylated might be further inhibited. A positive result would suggest that the enzymatic activity of proline hydroxylase was the limiting factor in the incomplete hydroxylation of proline at these positions. The data indicated that the hydroxyproline content of rat skin α1-CB2 from iron-deficient rats did not change. These results do not permit any conclusion regarding the reason for the incomplete hydroxylation of proline in rat collagen but they do exclude covert iron deficiency as an explanation for the findings reported here.

The elucidation of the sequence of α1-CB2 has provided a new opportunity to analyze the specificity of bacterial collagenase. The first studies of the specificity of collagenase were based on the analysis of enzymatic fragments obtained by digestion of collagen or the denatured protein, gelatin. These studies together with later experiments based on the action of collagenase on synthetic peptides (see Nordwig (1962) for review) indicated that collagenase cleaved the sequence X-Pro-Y-Gly-Pro-Z at the Y–Gly bond. In synthetic peptides

proline could be replaced by hydroxyproline and X, Y, and Z could be any amino acid. In addition, several peptides in which no imino acid was present on the amino side of the susceptible peptide bond or whose sequence could be characterized as X-Pro-Y-Gly-Z-Pro were also cleaved (Heyns and Legler, 1960; Nagai *et al.*, 1960; Poroshin *et al.*, 1961).

Table III summarizes the amino acid sequences surrounding the bonds which were cleaved by collagenase as well as the sequences around those bonds which were not susceptible to the enzyme. Of the 11 bonds joining the 12 triplets in α1-CB2, 7 were cleaved by collagenase. The susceptible bonds could be divided into two groups. The first group, of which there were five examples (Table III, Ia), was characterized by the sequence Gly-X-Y-Gly-Pro-Z. In one instance X was proline and in two instances Y was predominantly hydroxyproline but in two cases neither X or Y was an imino acid. Two bonds were split in which the surrounding amino acid sequence was characterized as Gly-X-Y-Gly-Z-Hyp (Table III, Ib). In both instances X was proline so that it is not known whether this is a requirement.

Under the conditions used for collagenase digestion evidence for partial cleavage was found in the case of only two bonds. While both of these were Ser–Gly bonds (Figure 1, residues 3–4 and 33–34) it seems more likely that the proximity of these residues to the NH_2- and COOH-terminal ends of α1-CB2 accounted for the relative resistance of the bonds to the enzyme.

The four bonds which resisted cleavage by collagenase could also be divided into two groups. In two instances (Table III, IIa) no imino acid was present in the triplet on the carboxy side of the potentially susceptible bond. In the other two instances (Table III, IIb) glutamic acid was adjacent to glycine in this triplet. It seems probable that the presence of the charged glutamyl side chain near the potentially susceptible bond inhibits the enzyme. Heyns and Legler (1960) have shown that the rate of hydrolysis by collagenase of the peptide CBz-Pro-Glu-Gly-Pro-NH_2 is significantly lower than the rates with peptides in which proline, alanine, or arginine are substituted for glutamic acid. The presence of glutamic acid on the carboxy side of glycine may have an even more profound inhibitory effect since it appears to be the composition of this triplet which primarily directs the specificity of collagenase.

TABLE III: The Specificity of Bacterial Collagenase as Deduced from an Analysis of the Products of Cleavage of α1-CB2.[a]

I. Bonds Cleaved
a. Gly-X-Y....Gly-Pro-Z
H_2-Gly-Pro-Ser....Gly-Pro-Arg– (1, 6)
-Gly-Leu-Hyp*...Gly-Pro-Hyp*- (7, 12)
-Gly-Ala-Hyp*...Gly-Pro-Gln- (13, 18)
-Gly-Phe-Gln....Gly-Pro-Hyp*- (10, 24)
-Gly-Ala-Ser....Gly-Pro-Hse-OH (13, 36)
b. Gly-X-Y....Gly-Z-Hyp
-Gly-Pro-Arg....Gly-Leu-Hyp*- (4, 9)
-Gly-Pro-Hyp*...Gly-Ala-Hyp*- (10, 15)
II. Sequences Resistant to Cleavage
a.
-Gly-Pro-Gln-Gly-Phe-Gln- (16, 21)
-Gly-Glu-Hyp*-Gly-Ala-Ser- (28, 33)
b.
-Gly-Pro-Hyp*-Gly-Glu-Hyp*- (22, 27)
-Gly-Glu-Hyp*-Gly-Glu-Hyp*- (25, 30)

[a] An asterisk indicates that proline can substitute for hydroxyproline. The numbers refer to residue numbers in Figure 1.

Acknowledgment

I am indebted to Dr. Karl A. Piez for his generous support, advice, and stimulating discussions, and to Dr. Andrew H. Kang for his assistance during the early phases of this work. I also wish to thank Mrs. Suzanne Roniger for her able technical assistance.

References

Blackburn, S. (1965), *Methods Biochem. Anal. 13*, 1.
Bornstein, P. (1967), *J. Biol. Chem. 242*, 2572.
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. (1964), *J. Clin. Invest. 43*, 1813.
Bornstein, P., and Piez, K. A. (1966), *Biochemistry 5*, 3460.
Chvapil, M., and Kobrle, V. (1961), *Experientia 17*, 226.
Deyl, Z., and Rosmus, J. (1965), *J. Chromatog. 20*, 514.
Franzblau, C., Gallop, P. M., and Seifter, S. (1963), *Biopolymers 1*, 79.
Gallop, P. M., and Seifter, S. (1962), *in* Collagen, Ramanathan, N., Ed., New York, N. Y., Interscience, pp 249–261.
Gottlieb, A. A., Kaplan, A., and Udenfriend, S. (1966), *J. Biol. Chem. 241*, 1551.
Grassmann, W., Hanning, K., and Nordwig, A. (1963), *Z. Physiol. Chem. 333*, 154.
Greenberg, J., Fishman, L., and Levy, M. (1964), *Biochemistry 3*, 826.
Heyns, K., and Legler, G. (1960), *Z. Physiol. Chem. 321*, 184.
Hurych, J., and Chvapil, M. (1965), *Biochim. Biophys.*

Acta 97, 361.
Hutton, J. J., Tappel, A. L., and Udenfriend, S. (1967), *Arch. Biochem. Biophys. 118*, 231.
Juva, K., and Prockop, D. J. (1966), *J. Biol. Chem. 241*, 4419.
Kang, A. H., Bornstein, P., and Piez, K. A. (1967), *Biochemistry 6*, 788.
Kivirikko, K. I., and Prockop, D. J. (1967a), *Biochem. J. 102*, 432.
Kivirikko, K. I., and Prockop, D. J. (1967b), *Proc. Natl. Acad. Sci. U. S. 57*, 782.
Kroner, T. D., Tabroff, W., and McGarr, J. J. (1955), *J. Am. Chem. Soc. 77*, 3356.
Lukens, L. N. (1966), *Proc. Natl. Acad. Sci. U. S. 55*, 1235.
Manning, J. M., and Meister, A. (1966), *Biochemistry 5*, 1154.
McCall, M. G., Newman, G. E., O'Brian, J. R. P., Valberg, L. S., and Witts, L. J. (1962), *Brit. J. Nutr. 16*, 297.
Miller, E. J., and Piez, K. A. (1966), *Anal. Biochem. 16*, 320.
Nagai, Y. (1961), *J. Biochem. (Tokyo) 50*, 486.
Nagai, Y., Sakakibara, S., Noda, H., and Akabori, S. (1960), *Biochim. Biophys. Acta 37*, 567.
Nordwig, A. (1962), *Leder 1*, 10.
Piez, K. A., Eigner, E. A., and Lewis, M. S. (1963), *Biochemistry 2*, 58.
Pine, E. K., and Holland, J. F. (1966), *Arch. Biochem. Biophys. 115*, 95.
Poroshin, K. T., Kozarenko, T. D., Shibnev, V. A., and Debabov, V. G. (1961), *Biokhimiya 26*, 219.
Ramachandran, G. N. (1963), *in* International Review of Connective Tissue Research, Vol. I, Hall, D. A., Ed., New York, N. Y., Academic, pp 127–182.
Rao, N. V., and Harrington, W. F. (1966), *J. Mol. Biol. 21*, 577.
Schroeder, W. A., Kay, L. M., LeGette, J., Honnen, L., and Green, F. C. (1954), *J. Am. Chem. Soc. 76*, 3556.
Schrohenloher, R. E., Ogle, J. D., and Logan, M. A. (1959), *J. Biol. Chem. 234*, 58.
Seifter, S., and Gallop, P. M. (1962), *Methods Enzymol. 5*, 659.
Seifter, S., Gallop, P. M., Klein, L., and Meilman, E. (1959), *J. Biol. Chem. 234*, 285.
Seiler, N., and Wiechmann, J. (1964), *Experientia 20*, 559.
Taniuchi, H., and Anfinsen, C. B. (1966), *J. Biol. Chem. 241*, 4366.
Udenfriend, S. (1966), *Science 152*, 1335.
Urivetzky, M., Frei, J. M., and Meilman, E. (1966), *Arch. Biochem. Biophys. 117*, 224.