Comparative Sequence Studies of Rat Skin and Tendon Collagen. II. The Absence of a Short Sequence at the Amino Terminus of the Skin α1 Chain*

Paul Bornstein

ABSTRACT: The amino acid sequence of rat tendon collagen is very similar but not identical with that of rat skin collagen. In addition to minor differences related to the hydroxylation of proline, the α1 chain of rat skin collagen lacks the tetrapeptide Glx-Met-Ser-Tyr which is present at the NH_2 terminus of the α1 chain of rat tendon collagen. This short sequence may be removed from the skin protein by a physiologic proteolytic mechanism or its existence may reflect the presence of different genes coding for tissue-specific collagen chains in the two tissues. However, limited proteolysis during extraction and purification of rat skin collagen cannot be excluded. If the primary structure and conformation at the NH_2 terminus of rat skin collagen differs *in vivo*, this modification could influence the aggregative properties and relative rates of cross-link formation of rat skin collagen *vis-à-vis* rat tendon collagen and play a role in collagen biogenesis. The possibility that several amino acids including glutamine (or pyrrolidone-5-carboxylic acid) are removed from the NH_2 terminus of a polypeptide chain is also of interest in view of the role which pyrrolidone-5-carboxylic acid may play in the initiation of polypeptide chain synthesis in higher organisms.

Collagen participates in a variety of structural complexes in different tissues of the same organism. The macromolecular organization of the protein varies in these tissues and reflects the particular function which collagen is required to fulfill. It is not known whether the fibrogenesis of collagen and its integration into a structural complex result at least in part from tissue-specific differences in the primary structure of the protein or whether mechanisms involving other macromolecules, such as glycosaminoglycans and glycoproteins, are responsible for these processes.

Relatively minor tissue-specific differences in the contents of hydroxyproline and hydroxylysine in collagen have been reported (Piez and Likins, 1957; Piez *et al.*, 1963; Fujimoto and Adams, 1964; Pine and Holland, 1965; Bornstein, 1967a,b). However, the significance of a variation in the degree of hydroxylation of these amino acids is not understood. Since proline and lysine are hydroxylated after their incorporation in peptide linkage (Udenfriend, 1966), the content of these amino acids in collagen probably results from the activity or nature of the tissue enzymes responsible for peptidyl hydroxylation of collagen rather than from different collagen genes.

The present study addresses itself to the question of whether other chemical differences exist in collagen and whether these can contribute to the tissue-specific properties of the protein. Although the α1 and α2 chains of rat skin collagen and rat tendon collagen are very similar as judged by chromatographic behavior and amino acid composition (Piez *et al.*, 1963), their large size (95,000 molecular weight) may obscure sequence differences between homologous chains. The α1 and α2 chains of rat tendon collagen were therefore subjected to a structural analysis employing cleavage of methionyl bonds with CNBr as performed in previous investigations of rat skin collagen (Bornstein and Piez, 1966; Kang *et al.*, 1967; Butler *et al.*, 1967). The α1 chain of rat tendon collagen and the homologous chain of rat skin collagen were found to be extremely similar in sequence but differed in the presence of an additional four amino acids at the NH_2 terminus of the rat tendon collagen chain. The possible explanations and implications of this finding are the main substance of this communication. As a consequence of this work additional evidence was found for the chromatographic heterogeneity of CNBr-produced peptides derived from the α1 chains of both rat skin collagen (Butler *et al.*, 1967) and rat tendon collagen. This heterogeneity is unrelated to the hydroxylation of proline or to the conversion of lysyl residues to lysyl-derived aldehydes (Bornstein and Piez, 1966; Piez *et al.*, 1966).

Materials and Methods

Preparation of Collagen. Rat skin and tail tendons were obtained from healthy 100–150-g male Sprague–Dawley rats and from rats made lathyritic by maintenance on a diet of 1.5 g of β-aminopropionitrile/kg of powdered pellet for 5 weeks. Salt (1.0 M NaCl) and 0.5 M acetic acid extracted collagens were extracted

* From the Department of Biochemistry and the Research Training Unit, Department of Medicine, University of Washington School of Medicine, Seattle, Washington 98105. *Received August 26, 1968.* Supported by National Institutes of Health Grant AM-11248 and National Science Foundation Grant GB-6079. A preliminary account of part of this work has appeared (Bornstein, 1967c).

and purified largely as described previously (Bornstein and Piez, 1966). Tissues were minced in cold 0.05 M Tris buffer (pH 7.5) in a Waring Blendor equipped with a Polytron attachment. All procedures employed in the preparation of 1 M NaCl-extracted collagen were performed at neutral pH, except for the final dialysis of the purified collagen for which 0.25% acetic acid was used. This precaution was taken to inhibit possible limited *in vitro* proteolysis of collagen by acid hydrolases which may be released during the initial stages of extraction of the protein.

Chromatography on CM-cellulose and Phosphocellulose. $\alpha 1$ and $\alpha 2$ chains and β_{12} components of rat skin collagen and rat tendon collagen were obtained by chromatography on CM-cellulose at pH 4.8 (Piez *et al.*, 1963; Bornstein and Piez, 1966). Chromatography of CNBr-produced fragments was performed on phosphocellulose (Bornstein and Piez, 1966) and on CM-cellulose at pH 4.8 (Bornstein and Piez, 1965) and pH 3.6 (Butler *et al.*, 1967). Under the conditions used for phosphocellulose chromatography only the smaller fragments, derived largely from the NH_2-terminal regions of the chain, were eluted. The use of 2.5 × 18 cm columns of phosphocellulose permitted the chromatography of 200–250 mg of CNBr-produced fragments at flow rates of 150–200 ml/hr. It was found convenient, at the conclusion of each run, to equilibrate the phosphocellulose with 0.3 N sodium formate (pH 3.8) followed by starting buffer (K. A. Piez, personal communication).

Molecular Sieve Chromatography. α chains, β components, and large CNBr-produced fragments were separated from buffer salts by chromatography on Sephadex G-25, coarse beads (Pharmacia, Inc.) (Piez *et al.*, 1963). Small CNBr-produced fragments were desalted by chromatography on Bio-Gel P-2 (100–200 or 200–400 mesh) (Bio-Rad Laboratories) equilibrated with 0.03 M ammonium propionate buffer (pH 4.5).

Cleavage with CNBr. CNBr (Eastman Organic Chemicals) was used without further purification. The technique employed has been described (Bornstein and Piez, 1965, 1966).

Enzymic Hydrolysis. Digestion with α-chymotrypsin (three-times crystallized, Worthington) and trypsin (two-times crystallized, Worthington) was performed in 0.2 M NH_4HCO_3 (pH 7.8) containing 1×10^{-3} M $CaCl_2$. The enzymes (1% of the substrate by weight) were added as a 0.5% solution in 1×10^{-3} M HCl. Digestion was limited to 1–2 hr at room temperature and was terminated by lyophilization.

Limited Cleavage of Native Collagen. Cleavage of native lathyritic rat skin collagen and rat tendon collagen with trypsin was performed at 20° for 4 hr as described previously (Bornstein *et al.*, 1966). The dialysates were lyophilized and chromatographed on Bio-Gel P-2 (200–400 mesh).

Edman Degradation, Dansylation, and Identification of Dimethylaminonaphthalene-5-sulfonyl Amino Acids. Amino acid sequences of isolated peptides were determined by sequential Edman degradation and identification of the original and successively newly formed NH_2-terminal amino acids by dansylation (Gray, 1967). The solution of peptide (5–50 mμmoles) in water was divided into the number of aliquots required (one for each step of the proposed degradation) and pipetted into conical glass-stoppered centrifuge tubes. The peptide was dried at 60° for 15 min in a heated vacuum desiccator. Coupling was performed under nitrogen at 45° for 1 hr in 100 μl of 5% phenyl isothiocyanate (distilled) in pyridine (refluxed with ninhydrin and distilled) diluted with 200 μl of 50% pyridine. The tubes were dried *in vacuo* at 60° for 30 min and the peptide was cleaved under nitrogen at 45° with 200 μl of trifluoroacetic acid for 30 min. The trifluoroacetic acid was removed by vacuum desiccation, 150 μl of H_2O was added, and extraction of diphenylthiourea and other by-products of the reaction was carried out three times with 1.5 ml of butyl acetate. The peptide, shortened by one residue at its NH_2-terminal end, was then transferred to a 6 × 50 mm test tube with 0.2 M $NaHCO_3$ and dried in a heated vacuum desiccator containing H_2SO_4. $NaHCO_3$ (10 μl of 0.2 M) and dimethylaminonaphthalene-5-sulfonyl chloride (10 μl) (recrystallized from isooctane, 2.5 mg/ml in acetone) were added. The tube was covered with parafilm and incubated at 37° for 1 hr. After drying, the dansylated peptide was hydrolyzed under nitrogen at 110° for 16 hr with 50 μl of doubly distilled constant-boiling HCl. If proline was suspected as NH_2 terminal, hydrolysis was limited to 8 hr. The HCl was removed by vacuum distillation and the hydrolysate was dissolved in a small quantity of 50% pyridine for application to a thin-layer plate.

Dimethylaminonaphthalene-5-sulfonyl amino acids were identified by thin-layer chromatography on silica gel plates or polyamide sheets (Woods and Wang, 1967). Unequivocal identification of each new NH_2-terminal amino acid was achieved by chromatography with two or more solvent systems. The following systems were used frequently: (a) methyl acetate–isopropyl alcohol–NH_3 (9:7:4) (Seiler and Wiechmann, 1967), (b) ethyl acetate–chloroform–methanol–acetic acid (10:10:1:1) (Crowshaw *et al.*, 1968), (c) H_2O–formic acid (100:1.5) (d) benzene–acetic acid (9:1) (Woods and Wang, 1967), and (e) chloroform–benzyl alcohol–acetic acid (70:30:3) (Deyl and Rosmus, 1965). Systems a and e were used primarily with silica gel plates and systems c and d with polyamide sheets. System b proved satisfactory with both media.

Spectrophotometric Detection of Aldehydes. Peptides were treated with *N*-methylbenzothiazolone hydrazone (Eastman Organic Chemicals) according to the method of Paz *et al.* (1965). Ultraviolet spectra were obtained using a Cary Model 14 recording spectrophotometer.

Peptide Maps. Two-dimensional chromatography and high-voltage electrophoresis of the tryptic digests of peptides $\alpha 1$-CB3 and $\alpha 2$-CB2 were performed as described by Butler *et al.* (1967).

Amino acid analyses were performed on a Beckman 120 C analyzer modified for accelerated single-column gradient elution largely according to the specifications described by Miller and Piez (1966). Using temperature programming and a slightly modified gradient both homoserine and homoserine lactone were adequately

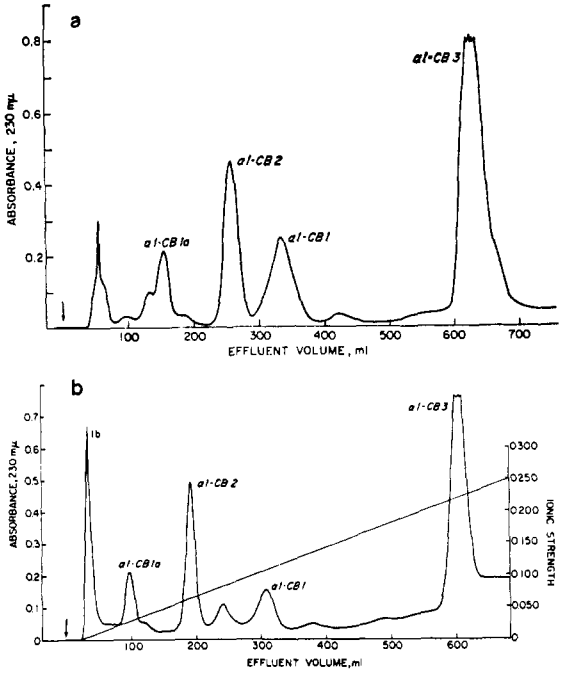

FIGURE 1: Elution patterns of peptides. (a) Obtained by CNBr cleavage of the α1 chain of acid-extracted rat tendon collagen. Chromatography was performed on phosphocellulose at pH 3.8, 40°. The arrow indicates placement of the sample (120 mg) dissolved in 10 ml of starting buffer, 0.001 M sodium acetate (pH 3.8). The shoulder on the ascending slope of peptide α1-CBla represents the peptide (α1-CBO)-(α1-CB1). (b) Phosphocellulose elution pattern of peptides obtained by CNBr cleavage of the α1 chain (120 mg) of acid-extracted rat skin collagen. The straight line plot indicates the ionic strength gradient. Reproduced with permission from *Biochemistry 5*, 3460 (1966).

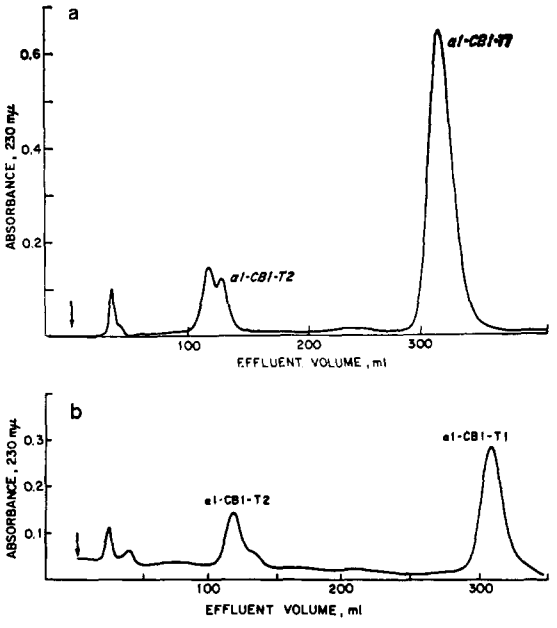

FIGURE 2: Phosphocellulose elution pattern of tryptic digests: (a) Of rat tendon collagen α1-CB1. Chromatography was performed at pH 3.8, 40°. The arrow indicates placement of the sample (1.2 μmoles) dissolved in 10 ml of starting buffer. The double peak of α1-CB1-T2 probably results from the equilibrium of homoserine with its lactone. (b) Of rat skin collagen α1-CB1 (1.0 μmole). See legend to Figure 2a for further details. Reproduced with permission from *Biochemistry 6*, 788 (1967).

separated from adjacent amino acids. Peptides were hydrolyzed under nitrogen in doubly distilled constant-boiling HCl at 110° for 24 hr. The following corrections were made for hydrolytic losses: threonine, 2.4%; serine, 6.2%; and methionine and tyrosine, 5%. A correction of 6.9% was applied for incomplete release of valine.

Results

The similarity of the α1 and α2 chains of rat skin collagen and rat tendon collagen, first suggested by their amino acid compositions (Piez *et al.*, 1963), is further supported by the present experiments which indicate a marked similarity in the CNBr-produced peptides of the two proteins. CM-cellulose chromatograms of CNBr digests of rat tendon collagen α1 and α2 chains at pH 4.8, and of the rat tendon collagen α1 chain at pH 3.6, were identical, within experimental error, with those previously obtained for rat skin collagen. However, subtle but significant differences were detected in some fragments as a result of their elution pattern from phosphocellulose.

The NH_2 Terminus of the α1 Chain. The phosphocellulose chromatogram of the CNBr digest of the α1 chain of rat tendon collagen is illustrated in Figure 1a. For comparison, a chromatogram of a similar digest from rat skin collagen is reproduced in Figure 1b. Certain differences such as the precise position of elution of individual peptides and the size of the initial or breakthrough peak are not reproducible and are of no apparent significance. The overlapping peptide resulting from lack of cleavage of the methionyl residue between α1-CB1 and α1-CB2 which elutes between the two peptides in Figure 1b forms the leading shoulder of the α1-CB1 peak in Figure 1a.[1] Two differences between Figures 1a,b were consistent and of significance. (1) A leading shoulder on the α1-CB1a peak in Figure 1a is absent in the corresponding location in Figure 1b. (2) The absorbance at 230 mμ of the α1-CB1 and α1-CB1a peaks *vis-à-vis* α1-CB2 is greater in rat tendon collagen (Figure 1a) than in rat skin collagen (Figure 1b).

On the basis of position of elution from phosphocellulose and amino acid composition, peptides α1-CB1 and α1-CB1a in rat tendon collagen are clearly homologous to the corresponding peptides from rat skin collagen. However, the amino acid composition of rat tendon collagen α1-CB1 (Table I) and α1-CB1a indicated the presence of an additional seryl and tyrosyl residue. As in the case of rat skin collagen, the amino acid composition of rat tendon collagen α1-CB1a differed from that of

[1] α1-CB1 and α1-CB2 are adjacent peptides on the α1 chain (Bornstein and Piez, 1966). The evidence related to the ordering of the other CNBr-produced peptides of the α1 chain of rat skin collagen has recently been summarized (Piez *et al.*, 1968).

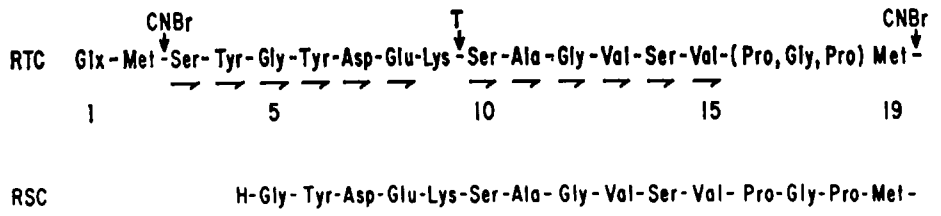

FIGURE 3: Proposed amino acid sequence for the NH_2 terminus of the $\alpha 1$ chain of rat tendon collagen. The sequence of the NH_2 terminus of the $\alpha 1$ chain of rat skin collagen, determined by Kang *et al.* (1967), is shown for comparison. The vertical arrows indicate points of cleavage by CNBr and trypsin (T). The horizontal arrows indicate the sequence determined by Edman degradation. Glx indicates that the residue is either glutamic acid, glutamine, or pyrrolidonecarboxylic acid.

TABLE I: Amino Acid Composition of NH_2-Terminal Peptides from Rat Tendon Collagen and Rat Skin Collagen.[a]

	Rat Tendon Collagen						Rat Skin Collagen[b]			
		α1-CB1					α1-CB1			
	α1-CB0		T1	T2	(α1-CB0)-(α1-CB1)	α2-CB1		T1	T2	α2-CB1
Aspartic acid	(0.1)	1.0	1.0		1.1	1.0	1.0	1.0		1.0
Serine	(0.1)	2.9	0.9	2.0	3.0	1.9	2.0		2.0	2.0
Homoserine[c]	0.8	1.0		0.9	1.2	1.0	0.9		0.9	1.0
Glutamic acid	1.1	1.1	1.0		1.9	1.0	1.0	1.0		1.0
Proline		1.9		1.9	2.2	1.9	1.8		1.8	1.9
Glycine	(0.2)	3.1	1.0	2.1	3.6	3.2	3.2	1.1	2.1	3.1
Alanine	(0.1)	1.0		1.1	1.1	1.1	1.1		1.0	1.0
Valine	(0.1)	2.1		2.0	1.8	1.1	1.8		2.1	1.0
Methionine					0.7					
Tyrosine		1.9	2.1		2.2	0.8	1.2	0.9		1.1
Lysine		1.0	0.9		0.6	0.9	1.0	1.0		0.9

[a] Values are given as residues per peptide. A space indicates that the amino acid was either entirely absent or present as less than 0.1 residue. Residues in parentheses are fractional residues thought to be impurities. [b] Obtained from Bornstein and Piez (1966). [c] Includes homoserine lactone.

α1-CB1 only in lacking the single lysyl residue. The evidence that α1-CB1a is derived from an α chain in which a lysyl residue has undergone oxidation deamination to a lysyl-derived aldehyde has been presented (Bornstein and Piez, 1966; Piez *et al.*, 1966).

Tryptic cleavage of α1-CB1 resulted in two peptides which were readily separable either by phosphocellulose or by Bio-Gel P-2 chromatography. The phosphocellulose elution patterns of the tryptic peptides of α1-CB1 from rat tendon collagen and rat skin collagen are illustrated in Figure 2a,b. The more basic lysine-containing peptides (α1-CB1-T1) also contain tyrosine (Table I) and have a distinctly higher extinction at 230 mμ. The additional tyrosyl residue in rat tendon collagen α1-CB1-T1 further increases the extinction of this peptide in comparison with α1-CB1-T2 and explains the higher absorbance of α1-CB1 and α1-CB1a from rat tendon collagen as compared with the corresponding peptides from rat skin collagen (Figure 1a,b).

The sequence of the first six amino acids in both rat tendon collagen α1-CB1-T1 and α1-CB1-T2 was determined by sequential dansylation and Edman degradation. These experiments indicated that the initial sequence in rat tendon collagen α1-CB1 is Ser-Tyr (residues 3–4, Figure 3) and confirmed the identity of the remainder of the peptide, insofar as determined, with rat skin collagen α1-CB1.

The amino acid composition of the peptide eluting just before rat tendon collagen α1-CB1a (Figure 1a) indicated that this peptide ((α-CB0)-(α1-CB1), Table I) contains an additional two amino acids, glutamic acid (or glutamine) and methionine. The presence of a methionyl-containing peptide in small amounts is not unexpected since only 90–92% of methionyl bonds were cleaved with CNBr under the conditions used.[2]

Further evidence for the sequence Glx-Met[3] in the

[2] In some instances modification of methionyl residues to homoserine (or a product which yields homoserine after acid hydrolysis) occurred without concomitant cleavage of the peptide chain. The amino acid composition of the peptide eluting after α1-CB1 in both Figure 1a and 1b suggests that it consists of α1-CB1 linked to α1-CB2 by a homoseryl residue. It is possible that the iminolactone formed as an intermediate during CNBr cleavage (Gross and Witkop, 1962) is not completely hydrolyzed in 0.1 N HCl.

[3] Glx indicates that a distinction has not been made between glutamine, glutamic acid, and pyrrolidone-5-carboxylic acid.

α1 chain of rat tendon collagen was obtained by analysis of the peak eluting at the start of the phosphocellulose chromatogram of a CNBr digest of rat tendon collagen α1 chain. This material was chromatographed on Bio-Gel P-2 (100–200 mesh) and was found to contain a small peak, incompletely resolved from buffer salts, which yielded largely glutamic acid and homoserine (lactone) after acid hydrolysis (α1-CB0, Table I).[4] Attempts to detect a free NH_2-terminal group on the dipeptide (by dansylation) were unsuccessful.

These data indicate that the NH_2-terminal sequence of the rat tendon collagen α1 chain is Glx-Met-Ser-Tyr-. The dipeptide Glx-Met must be located on the amino side of α1-CB1 since the overlapping peptide (α1-CB1)-(α1-CB2) resulting from incomplete cleavage of the second methionyl residue in the α1 chain of rat tendon collagen (residue 19, Figure 3) lacks the additional two amino acids.

To confirm the presence of the additional sequence at the NH_2 terminus of the α1 chain of rat tendon collagen, native lathyritic rat tendon collagen was subjected to limited cleavage with trypsin and the resulting dialyzable peptides were chromatographed on Bio-Gel P-2 (200–400 mesh). As shown previously for rat skin collagen (Bornstein *et al.*, 1966) these peptides are derived in large part from the NH_2-terminal ends of the α chains. This region lacks the helical conformation of the remainder of the collagen molecule and is therefore selectively susceptible to tryptic attack under the conditions used for digestion. Lathyritic collagen is required for this experiment since the oxidative deamination which normally occurs at the first lysyl residue in the chain prevents tryptic cleavage at this position unless this transformation is inhibited by the administration of lathyrogens.

The chromatograms in Figure 4a,b compare the elution patterns of the dialyzable peptides obtained from rat tendon collagen and rat skin collagen. The peak eluting at approximately 110–125 ml of effluent volume in both chromatograms represents a mixture of two pentapeptides which, in the case of rat tendon collagen, consists of residues 5–9 of the α1 chain and the homologous peptide from the α2 chain. An additional peak, eluting at 145–150 ml of effluent volume, was present in the preparation obtained from rat tendon collagen. Although heterogeneous, this peak contained predominantly serine, glutamic acid, methionine, and tyrosine in equivalent amounts after acid hydrolysis. This composition is consistent with the existence of the sequence Glx-Met-Ser-Tyr at the NH_2 terminus of the α1 chain of rat tendon collagen. The presence of this peptide in a tryptic digest is unexpected, but probably resulted from chymotrypsin-like activity in the commercial trypsin preparation.

[4] α1-CB1 had previously been identified as the NH_2-terminal peptide in the α1 chain of rat skin collagen (Bornstein *et al.*, 1966). In view of the almost certain homology of the rat tendon collagen α1 chain, the nomenclature established for the peptides derived from rat skin collagen will be applied to the corresponding peptides from rat tendon collagen. The additional dipeptide in the rat tendon collagen α1 chain is therefore designated α1-CB0.

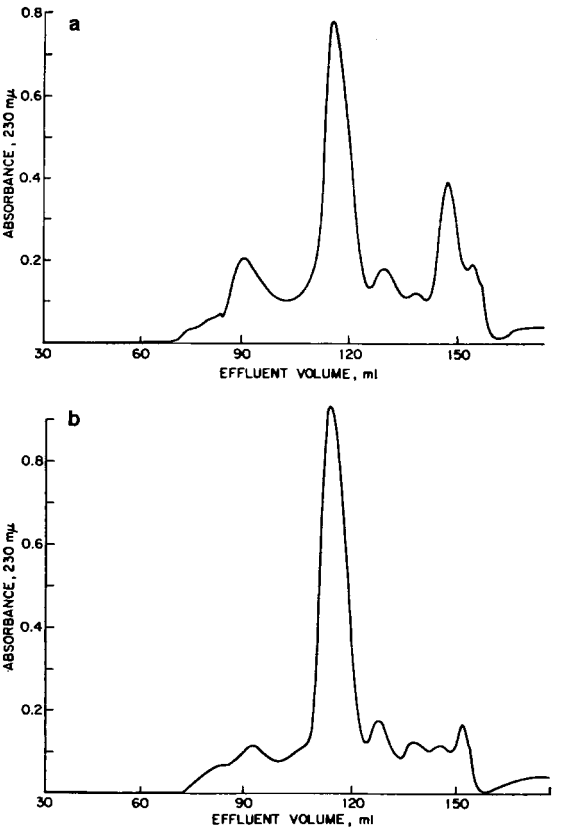

FIGURE 4: Elution pattern of dialyzable peptides. (a) Obtained by limited tryptic digestion of native lathyritic rat tendon collagen (250 mg). Chromatography was performed on Bio-Gel P-2 (200–400 mesh) equilibrated with 0.03 M ammonium propionate buffer (pH 4.5). (b) Obtained by limited tryptic digestion of native lathyritic rat skin collagen (250 mg). Chromatography was performed on Bio-Gel P-2 (200–400 mesh) equilibrated with 0.03 M ammonium propionate buffer (pH 4.5).

The Presence of Aldehydes at the NH_2 Terminus of Rat Tendon Collagen. The absorption spectra of the azine derivatives of α1-CB1a and α2-CB1a with *N*-methylbenzothiazolone hydrazone indicated the presence of a saturated aldehyde in these peptides whereas the cross-linked peptide, β_{12}-CB1, contained largely an α,β-unsaturated aldehyde. The data are analogous to those obtained with rat skin collagen (Bornstein and Piez, 1966) and indicate that the mechanism of formation of intramolecular cross-links in tendon collagen is similar to that in the skin protein.

The NH_2 Terminus of the α2 Chain. The phosphocellulose chromatogram of the CNBr digest of the α2 chain of rat tendon collagen was identical, within experimental error, with the corresponding chromatogram from rat skin collagen. The amino acid composition of α2-CB1 from rat tendon collagen was also identical with that of the homologous peptide from rat skin collagen (Table I). The lysyl-derived aldehyde-containing peptide, α2-CB1a, is present in tendon and, as in skin, differs from α2-CB1 only in lacking the single lysyl residue.

In view of the suggestion that lysyl-derived aldehyde

TABLE II: Amino Acid Composition of Two Peptides from the Helical Region of Rat Tendon Collagen and Rat Skin Collagen.[a]

	Rat Tendon Collagen α1-CB3		Rat Skin Collagen α1-CB3		Rat Tendon Collagen α2-CB2	Rat Skin Collagen α2-CB2
	Residues/ 1000	Residues/ Peptide	Residues/ 1000	Residues/ Peptide	Residues/ Peptide	Residues/ Peptide
3-Hydroxyproline					0.1	
4-Hydroxyproline	103	15.4	102	15.2	2.7	2.8
Aspartic acid	44.8	6.7	47.0	7.0	2.9	2.8
Threonine	12.5	1.9	14.1	2.1	1.0	1.0
Serine	19.2	2.9	19.8	3.0	1.0	1.1
Homoserine[b]	5.9	0.9	6.1	0.9	0.9	0.9
Glutamic acid	107	16.0	108	16.1	1.1	1.1
Proline	95.6	14.3	92.4	13.8	3.1	3.1
Glycine	330	49.3	336	50.0	9.8	9.4
Alanine	137	20.4	135	20.1	2.2	2.1
Valine	28.5	4.3	28.0	4.2	1.0	1.0
Leucine	20.3	3.0	20.0	3.0	1.0	1.0
Phenylalanine	19.8	3.0	18.9	2.8		
Hydroxylysine	1.1	0.2	0.8	0.1		
Lysine	32.9	4.9	31.1	4.6		
Arginine	42.1	6.3	39.9	5.9	3.0	3.0

[a] A space indicates that the amino acid was either entirely absent or present as less than 0.1 residue. [b] Includes homoserine lactone.

formation in the α2 chain of collagen is not inhibited by administration of lathyrogens (Rojkind and Juarez, 1966) the relative quantities of α2-CB1 and α2-CB1a in CNBr digests of α2 from lathyritic and normal animals were determined. Amino acid analyses of the relevant peaks in phosphocellulose chromatograms indicated that 79% of the NH_2-terminal peptides in preparations of α2 from lathyritic rat tendon collagen was in the lysine (CB1) form. In contrast, preparations of α2 derived from normal rat tendon collagen contained 80% or more of the NH_2-terminal peptides in the aldehyde (CB1a) form. These data indicate that β-aminopropionitrile fumarate inhibits the oxidative deamination of susceptible lysyl residues in both the α1 and α2 chains of collagen.

Peptides from the Helical Regions of the α1 and α2 Chains. The amino acid composition and sequence of α1-CB2 from rat skin collagen and rat tendon collagen have been reported (Bornstein, 1967a,b). These studies indicated that the sequences are identical with the exception of the degree of hydroxylation of proline. The amino acid compositions of two additional peptides are listed in Table II and are compared with the homologous peptides from rat skin collagen. As judged by their amino acid compositions, these peptides, as well as α1-CB2, are able to assume a triple-helical conformation and differ in this respect from the NH_2-terminal regions of the α chains. No significant differences in composition, apart from the presence of 3-hydroxyproline in rat tendon collagen α2-CB2, are apparent. The identity in the basic amino acid sequences of the peptides from the two tissues is further suggested by the marked similarity of the peptide maps obtained after tryptic hydrolysis. However, the α1-CB3 fraction from both tendon and skin collagen, when chromatographed on CM-cellulose, was resolved into a number of components (Figure 5). All four components were found to have the same or very similar amino acid compositions. Elevation of the pH to 10.7 for several minutes prior to chromatography (Butler *et al.*, 1967) indicated that α1-CB3 and α1-CB3''' are different chromatographic forms of the same peptide, resulting from the equilibrium between homoserine and its lactone. A similar relation exists between α1-CB3' and α1-CB3''. The reasons for the chromatographic separation of α1-CB3 and α1-CB3' (or α1-CB3'' and α1-CB3''') remain unexplained but the existence of these chromatographic forms suggests that the microheterogeneity of collagen extends to the presence, within the same tissue, of minor collagen components differing in basic structure from the predominant molecular species.

Discussion

Tissue-specific differences in protein structure and function have been recognized in the isoenzymes and have contributed to an understanding of the mode of action of some of these enzymes (Kaplan, 1965; Rutter *et al.*, 1968). The macromolecular structure and

the function of collagen vary widely in different tissues of the same species. It is not known whether the dissimilar properties of collagen in tissues such as tendon and skin are a consequence of differences which are intrinsic to the proteins synthesized in these tissues or whether the manner in which the proteins are organized within structural complexes results largely from the participation of other macromolecular mechanisms (Jackson and Bentley, 1968). The resolution of this question is of importance to an understanding of morphogenesis and may indicate the extent to which acquired disorders can be expected to contribute to deranged structure and function in connective tissues.

In order to interpret comparative sequence studies of collagen, it is first necessary to evaluate the heterogeneity of the protein and to determine whether the nature and degree of this heterogeneity differs from tissue to tissue. A known source of heterogeneity in both rat skin collagen and rat tendon collagen relates to the incomplete hydroxylation of individual prolyl residues (Bornstein, 1967a,b). The degree of hydroxylation in skin was found to exceed that in tendon in the region of the α1 chain represented by α1-CB2. In the present studies 0.1 equiv/peptide of 3-hydroxyproline was found in rat tendon collagen α2-CB2, whereas this imino acid was not detected in the homologous peptide from rat skin collagen. Although a role for 4-hydroxyproline in the formation of intermolecular hydrogen bonds has been suggested (Fujimoto, 1968) the significance and function of the two imino acids in collagen require further investigation. It is likely that the differences in hydroxylation at both positions 3 and 4 of proline result from variations in the nature or activity of the hydroxylating systems in skin and tendon rather than from differences in the sequence of the protocollagen synthesized in the two tissues.

The presence of fractional residues of hydroxylysine in α1-CB3 from both rat skin collagen and rat tendon collagen (Table II) suggests that incomplete hydroxylation of lysine also occurs. Partial hydroxylation of lysine in collagen has also been reported by Butler (1968). Since disaccharides in collagen are attached by *O*-glycosidic bonds to hydroxylysine (Butler and Cunningham, 1966; Spiro, 1967) tissue-specific variations in the carbohydrate content of collagen chains may be found.

The existence of at least two chromatographic forms of α1-CB3 noted in this work together with the previously observed heterogeneity of other CNBr-produced collagen peptides fractionated on CM-cellulose (Bornstein and Piez, 1965; Butler *et al.*, 1967) suggest that minor collagen components differing in basic structure exist. Such components could provide an additional source of heterogeneity in collagen, but this polymorphism appears not to be related to tissue-specific differences in the protein.

The present studies indicate that the α1 chain of rat tendon collagen differs from the homologous chain of rat skin collagen in containing an additional four amino acids at its NH_2-terminal end. It is possible that the genetic information required for the synthesis of the additional tetrapeptide in the tendon protein has

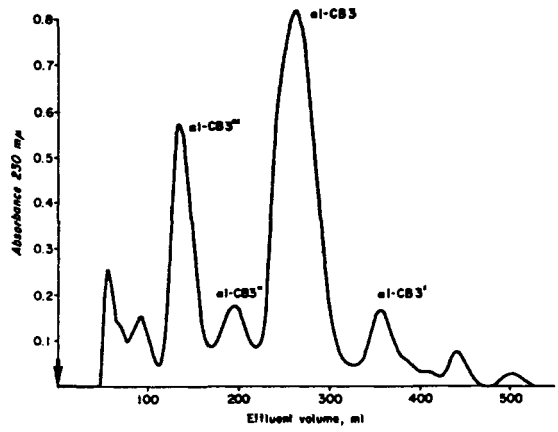

FIGURE 5: CM-cellulose elution pattern of an α1-CB3 fraction obtained by phosphocellulose chromatography of a CNBr digest of rat tendon collagen α1 (see Figure 1a). The arrow indicates placement of the sample (40 mg) dissolved in 25 ml of starting buffer, 0.02 M sodium acetate (pH 4.8). A linear salt gradient from 0 to 0.1 M NaCl over a volume of 800 ml was employed.

been deleted in the cistron coding for the skin collagen. However, it seems more likely that the two chains are initially the same in this region. In addition to the similarity of the remainder of the chains, the presence of the same or homologous tetrapeptide in chicken skin and bone collagen (Piez *et al.*, 1968) and in human skin collagen (Bornstein, 1968) argues for its existence in rat skin collagen.

There are at least two other explanations for the absence of the tetrapeptide from the NH_2 terminus of the α1 chain of rat skin collagen. (1) This sequence may be cleaved *in vivo* by a physiologic proteolytic mechanism. (2) Limited degradation of the skin protein may occur during extraction and purification. Since skin is a more cellular tissue than tendon, limited proteolysis of collagen by tissue cathepsins may occur more readily in homogenates of this tissue. Indirect support for such *in vitro* degradation of collagen derives from the observation that on two occasions preparations of rat tendon collagen α1 chain lacked the NH_2-terminal tetrapeptide entirely. The CNBr-produced peptides from these preparations were identical with those from rat skin collagen α1. However, attempts to inhibit a postulated proteolytic activity during the extraction of collagen from skin, and therefore to retain the additional tetrapeptide, have not been successful.

The NH_2 terminus of the collagen molecule is involved in intramolecular cross-linking (Bornstein *et al.*, 1966; Bornstein and Piez, 1966) and may also participate in the formation of intermolecular bonds. Although the initial step in cross-link formation, the conversion of lysyl residues to lysyl-derived aldehydes (Bornstein and Piez, 1966), is undoubtedly catalyzed enzymatically (Pinnell *et al.*, 1968; Page and Benditt, 1967), subsequent steps may proceed spontaneously as a result of the appropriate apposition of reactive groups in the local environment created by the aggregation of collagen molecules. In this respect the nonhelical region

of collagen may be considered analogous to the active site of an enzyme. Removal of a small segment of the α chain at the NH_2 terminus could affect the conformation of this region and could influence both the aggregative properties of the molecule as well as the rate of formation of cross-links. Since a tyrosyl residue exists in the additional sequence present in rat tendon collagen α1, and since the nonhelical regions in collagen contribute significantly to the immunogenicity of the protein (Schmitt *et al.*, 1964; Davison *et al.*, 1967), antigenic differences between skin and tendon collagen may be found.

The additional tetrapeptide in the α1 chain of rat tendon collagen is also of interest because the initial residue may be pyrrolidone-5-carboxylic acid. Both the failure of α1-CB0 to react with dimethylaminonaphthalene-5-sulfonyl chloride and the position of elution from phosphocellulose of the overlapping peptide (α1-CB0)-(α1-CB1) (Figure 1a) suggest that these peptides lack a free α-amino group. However, the ease with which glutamine and glutamic acid cyclize under the conditions of low pH used during the isolation of collagen and CNBr cleavage precludes the conclusion that pyrrolidone-5-carboxylic acid represents the NH_2-terminal residue of rat tendon collagen α1 *in vivo*.

The presence of glutamine or pyrrolidone-5-carboxylic acid as the NH_2-terminal amino acid in a number of proteins has been discussed (Blombäck, 1967). The α2 chain of rat skin collagen is known to contain a blocked NH_2 terminus which was identified as pyrrolidone-5-carboxylic acid after CNBr degradation (Kang *et al.*, 1967). In view of the amino acid composition (Table I) and chromatographic behavior of α2-CB1, the NH_2-terminal sequence of the α2 chain of rat tendon collagen appears to be identical with the corresponding region of rat skin collagen α2. Preliminary studies of the structure of the α1 chain of human skin collagen (Bornstein, 1968) and of the α1 chains of chicken skin and bone collagen (Piez *et al.*, 1968) also indicate, on the basis of homology with the rat proteins, that glutamine or its cyclized derivative is the NH_2-terminal residue in these chains.

It has been suggested that cyclization of glutamic acid on a tRNA may serve in chain initiation in mammalian proteins in a manner analogous to the role of *N*-formylmethionyl-tRNA in *E. coli* (Moav and Harris, 1967). Deletion or cleavage of NH_2-terminal pyrrolidone-5-carboxylic acid is likely in the light chains of γ-globulin synthesized by some mouse plasma cell tumors (Appella and Perham, 1968) and in the light chains of some human Bence–Jones proteins (Putnam *et al.*, 1967). Recently Doolittle and Armentrout (1968) have partially purified and characterized an enzyme, pyrrolidonyl peptidase, which selectively removes NH_2-terminal pyrrolidone-5-carboxylic acid from polypeptide chains in a strain of *Pseudomonas fluorescens*. In rat skin a physiologic mechanism may exist for the selective removal of a short NH_2-terminal sequence from collagen and such a mechanism could play a role in the morphogenesis of the tissue. However, the existence of reactions leading to a modification of this nature do not exclude the participation of other macromolecular mechanisms in the tissue-specific biogenesis of collagen.

Acknowledgment

The early phases of this work were performed in the laboratory of Dr. Karl A. Piez whom I thank for support and encouragement. The skillful technical assistance of Felicia Arguelles and Carolyn Zetes is gratefully acknowledged.

References

Appella, E., and Perham, R. N. (1968), *J. Mol. Biol. 33*, 963.
Blombäck, B. (1967), *Methods Enzymol. 11*, 398.
Bornstein, P. (1967a), *J. Biol. Chem. 242*, 2572.
Bornstein, P. (1967b), *Biochemistry 6*, 3082.
Bornstein, P. (1967c), *Abstracts, 7th Intern. Congr. Biochem., Tokyo*, 602.
Bornstein, P. (1968), *Science 161*, 592.
Bornstein, P., Kang, A. H., and Piez, K. A. (1966), *Biochemistry 5*, 3803.
Bornstein, P., and Piez, K. A. (1965), *Science 148*, 1353.
Bornstein, P., and Piez, K. A. (1966), *Biochemistry 5*, 3460.
Butler, W. T. (1968), *Science 161*, 796.
Butler, W. T., and Cunningham, L. W. (1966), *J. Biol. Chem. 241*, 3882.
Butler, W. T., Piez, K. A., and Bornstein, P. (1967), *Biochemistry 6*, 3771.
Crowshaw, K., Jessup, S. J., and Ramwell, P. W. (1968), *Biochem. J. 103*, 79.
Davison, P. F., Levine, L., Drake, M. P., Rubin, A., and Bump, S. (1967), *J. Exptl. Med. 126*, 331.
Deyl, Z., and Rosmus, J. (1965), *J. Chromatog. 20*, 514.
Doolittle, R. F., and Armentrout, R. W. (1968), *Biochemistry 7*, 516.
Fujimoto, D. (1968), *Biochim. Biophys. Acta 154*, 183.
Fujimoto, D., and Adams, E. (1964), *Biochem. Biophys. Res. Commun. 17*, 437.
Gray, W. R. (1967), *Methods Enzymol. 11*, 139, 469.
Gross, E., and Witkop, B. (1962), *J. Biol. Chem. 237*, 1856.
Jackson, D. S., and Bentley, J. P. (1968), *in* Treatise on Collagen, Vol. 2, Gould, B. S., Ed., New York, N. Y., Academic, p 189.
Kang, A. H., Bornstein, P., and Piez, K. A. (1967), *Biochemistry 6*, 788.
Kaplan, N. O. (1965), *in* Evolving Genes and Proteins, Bryson, V. and Vogel, H. S., Ed., New York, N. Y., Academic, p 243.
Miller, E. J., and Piez, K. A. (1966), *Anal. Biochem. 16*, 320.
Moav, B., and Harris, T. N. (1967), *Biochem. Biophys. Res. Commun. 29*, 773.
Page, R. C., and Benditt, E. P. (1967), *Biochemistry 6*, 1142.
Paz, M. A., Blumenfeld, O. O., Rojkind, M., Henson, E., Furfine, C., and Gallop, P. M. (1965), *Arch. Biochem. Biophys. 109*. 548.
Piez, K. A., Bladen, H. A., Bornstein, P., Butler, W. I.,

Kang, A. H., Lane, J. M., and Miller, E. J. (1968), *Brookhaven Symp. Biol.* (in press).
Piez, K. A., Eigner, E. A., and Lewis, M. S. (1963), *Biochemistry 2*, 58.
Piez, K. A., and Likins, R. C. (1957), *J. Biol. Chem. 229*, 101.
Piez, K. A., Martin, G. R., Kang, A. H., and Bornstein, P. (1966), *Biochemistry 5*, 3813.
Pine, E. K., and Holland, J. F. (1965), *Arch. Biochem. Biophys. 115*, 95.
Pinnell, S., Martin, G. R., and Gee, (1968), *Fed. Proc. 27*, 588.
Putnam, F. W., Shinoda, T., Titani, K., and Wikler, M. (1967), *Science 157*, 1050.
Rojkind, M., and Juarez, H. (1966), *Biochem. Biophys. Res. Commun. 25*, 481.
Rutter, W. J., Rajkumar, T., Penhoet, E., Kochman, M., and Valentine, R. (1968), *Ann. N. Y. Acad. Sci. 151*, 102.
Schmitt, F. O., Levine, L., Drake, M. P., Rubin, A. L., Pfahl, D., and Davison, P. F. (1964), *Proc. Natl. Acad. Sci. U. S. 51*, 493.
Seiler, N., and Wiechmann, M. (1967), *J. Chromatog. 28*, 351.
Spiro, R. G. (1967), *J. Biol. Chem. 242*, 4813.
Udenfriend, S. (1966), *Science 152*, 1335.
Woods, K. R., and Wang, K.-T. (1967), *Biochim. Biophys. Acta 133*, 369.

Staphylococcal Enterotoxin C. I. Phenolic Hydroxyl Ionization*

Concordia R. Borja

ABSTRACT: The ionization of the phenolic groups of enterotoxin C produced by *Staphylococcus aureus* strain 137 has been studied by spectrophotometric titration at 295 mμ in the range between pH 7.5 and 13.0. Of the 21 tyrosyl residues per mole of protein, 5 residues are deduced to be "free," located on the surface of the molecule, freely exposed to the solvent and hence easily accessible to OH^-. The ionization of these five "free" tyrosyl groups is reversible with $pK_{app} = 10.02$ and with no time dependence. The remaining 16 tyrosyl residues are postulated to be embedded in the interior ("buried") and capable of ionizing only after unfolding of the protein molecule. The buried tyrosyl groups ionize at pH values higher than 11.0 and the ionization process is not reversible and is time dependent. Six of the buried tyrosyl residues ionize between pH 11.0 and 12.0 with a pK_{app} of 11.5 and ten ionize above pH 12.0. Treatment of enterotoxin C with 5 M guanidine hydrochloride results in the normalization of all the 21 tyrosyl groups with a pK_{app} of 9.9.

The existence of two types of tyrosyl residues, free and buried, has also been demonstrated by studying the reaction of enterotoxin C with N-acetylimidazole, tetranitromethane, and tyrosinase. Five tyrosyl residues per mole of protein react with *N*-acetylimidazole at pH 7.5, five to six residues react with tetranitromethane at pH 8.0, and five groups are oxidized by tyrosinase at pH 6.5.

Enterotoxin C elaborated by *Staphylococcus aureus* strain 137 is a protein with a molecular weight of 34,100 (Borja and Bergdoll, 1967), and consists of a single polypeptide chain cross-linked by one disulfide bridge (Huang *et al.*, 1967). Amino acid analysis of enterotoxin C revealed 21 tyrosyl residues per molecule (Huang *et al.*, 1967).

Since the pioneering work of Crammer and Neuberger (1943) on the ionization of the phenolic hydroxyl groups of tyrosine, insulin, and egg albumin, many investigators have adapted the spectrophotometric method to follow the dissociation of hydrogen ions from the phenolic groups of both native and modified proteins. Spectrophotometric titrations have revealed at least two types of tyrosyl residues in many proteins (Beaven and Holiday, 1952; Wetlaufer, 1962), namely, residues which instantaneously ionize in a manner similar to the free tyrosine molecule ("free" or "exposed") and residues which are either hydrogen bonded to specific acceptor groups or embedded in the protein interior and capable of ionizing only slowly or after denaturation ("buried").

The states of tyrosyl residues in proteins have been the object of many investigations employing spectrophotometric titration, iodination, solvent perturbation techniques, and use of specific reagents as *N*-acetylimidazole (Riordan *et al.*, 1965), tetranitromethane (Sokolovsky *et al.*, 1966), and cyanuric fluoride (Kurihara *et al.*, 1963). Tyrosinase has also been used by several investigators (Sizer, 1953; Yasunobu *et al.*, 1959; Frieden *et al.*, 1959; Lissitzky *et al.*, 1960; Cory and Frieden, 1967), to determine the accessibility of the tyrosyl groups on the surface of the protein and to gain insight as to the relationship of these tyrosyl groups to the ac-

* From the Food Research Institute, University of Wisconsin, Madison, Wisconsin 53706. *Received June 7, 1968.* This investigation was aided by a contract with the U. S. Army Biological Center, Fort Detrick, Frederick, Md.