Structure of Rat Skin Collagen α1-CB8. Amino Acid Sequence of the Hydroxylamine-Produced Fragment HA2†

Gary Balian, Eva Marie Click, Mark A. Hermodson, and Paul Bornstein*,‡

ABSTRACT: α1-CB8, a large peptide obtained by cyanogen bromide cleavage of the α1 chain of rat collagen, can be split with hydroxylamine at an asparaginyl-glycyl bond to yield two fragments, HA1 and HA2. The amino acid sequence of the 99 residue NH_2-terminal fragment, HA1, has been reported (*Biochemistry 10*, 4470, 1971). The sequence of the 180 residues in HA2, the COOH-terminal fragment, has now been determined. As in the other collagen sequences the repeating triplet Gly-X-Y extends throughout HA2. Except for one instance in which glycine also occurs in position X, glycine is limited to the first position in this triplet; hydroxyproline is restricted to position Y. Both in this, and in other established collagen sequences, phenylalanine and leucine are restricted to position X. The distribution of some other amino acids also shows a preference for either position X or Y. The clustering of charged residues in HA2 correlates well with the electron optical pattern of segment-long-spacing aggregates of collagen. Homology of the α1 and α2 chains is indicated by the finding that a 30 residue sequence in α1-CB8 shows a high degree of identity with α2-CB2, particularly in distribution of charged amino acids.

The study of the primary structure of α1-CB8, a peptide obtained by cyanogen bromide cleavage of the α1 chain of rat collagen (Butler *et al.*, 1967), was undertaken to elucidate several important aspects of collagen structure. Knowledge of a sequence of this length (279 amino acids) would provide additional information relating the structure of collagen to its function as a structural protein with fiber-forming properties. When coupled with sequence data of the α2 chain these studies may further delineate interactions which contribute to the stability of the triple-helical molecule. Finally, the distribution of amino acids in collagen may reveal sequence homologies either within the α1 chain or between α1 and α2 which could reflect duplication of genetic material and provide a clue to the evolution of this unusual protein.

α1-CB8 constitutes about 25% of the length of the α1 chain of rat collagen and is located in the first half of the chain (Rauterberg and Kühn, 1968; Piez *et al.*, 1969). The finding that hydroxylamine cleaved this peptide at about one-third of its length from the NH_2 terminus (Bornstein, 1969a) yielding two fragments, HA1 (99 residues) and HA2 (180 residues), greatly simplified the task of sequence determination. The hydroxylamine-sensitive bond was identified as a cyclic imide which formed as a result of condensation of an asparaginyl side chain with the adjacent glycyl amide group in the polypeptide chain (Bornstein, 1970).

The first two papers in this series described the fractionation and amino acid composition of the tryptic peptides from α1-CB8 (Bornstein, 1970) and the amino acid sequence of HA1 (Balian *et al.*, 1971). The primary structure of HA2 is reported in this communication.

Materials and Methods

Preparation of α1-CB8-HA2. α1-CB8 was prepared from salt-extracted rat skin collagen (Butler *et al.*, 1967; Bornstein, 1970). The cyanogen bromide peptide was cleaved with hydroxylamine (Bornstein, 1970; Balian *et al.*, 1971) and the products separated by molecular sieve chromatography on 8% agarose (Bio-Rad Laboratories) equilibrated with 1 M $CaCl_2$ containing 0.05 M Tris-HCl (pH 7.5) (Piez, 1968; Bornstein, 1970). HA2 was purified by CM-cellulose chromatography in 0.02 M sodium acetate (pH 4.8) using a linear gradient of NaCl from 0 to 0.1 M (Balian *et al.*, 1971).

Enzymatic Hydrolyses. Collagenase (CLSPA, Worthington) was purified further on Sephadex G-200 and used in the presence of the sulfydryl reagent *N*-ethylmaleimide (Peterkofsky and Diegelmann, 1971). Tryptic peptides from HA2 (Bornstein, 1970) were incubated with collagenase at 37° for 16 hr (Balian *et al.*, 1971).

Digestion of α1-CB8 with thermolysin (Daiwa Kasei, K. K. Osaka, Japan) was performed in 0.2 M NH_4HCO_3 at 37° using a 1:100 ratio of enzyme to substrate. For digestion with chymotrypsin, 35 mg of HA2 was dissolved in 3 ml of 0.2 M NH_4HCO_3 containing 10^{-3} M $CaCl_2$ and the pH was adusted to 8.0. α-Chymotrypsin (TLCK[1] treated, Worthington) was added to a final concentration of 1:100 (enzyme to substrate by weight) and incubated at 37°. After 3 hr an equal amount of enzyme was added and digestion continued for 15 hr. Digestions with papain, trypsin, and carboxypeptidase were performed as described previously (Bornstein, 1967, 1970; Balian *et al.*, 1971).

Ion-exchange chromatography on DC-1 (Durrum Chemical Corp.), Dowex 50-X4, and Dowex 1 and *paper electrophoresis* at pH 6.5 were performed as described previously (Bornstein, 1970; Balian *et al.*, 1971).

Determination of Amide Groups. The presence or absence of side-chain amides of aspartic and glutamic acids was determined by the electrophoretic mobility of peptides at pH 6.5 or by identification of the PTH derivative using a Beckman gas chromatograph (Pisano and Bronzert, 1969).

Amino Acid Sequence Determination. The amino acid sequence of the two small peptides T5 and T10 was determined by the dansyl-Edman procedure (Bornstein, 1969b). The

† From the Department of Biochemistry, and Research Training Unit, Department of Medicine, University of Washington, Seattle, Washington 98195. *Received May 26, 1972.* This work was supported by National Institutes of Health Contract 69-2230 and National Institutes of Health Grants AM 11248 and GM 1573104.

‡ Recipient of Research Career Development Award K4-AM-42582 from the U. S. Public Health Service.

[1] Abbreviations used are: PTH, phenylthiohydantoin; TLCK, tosyllysine chloromethyl ketone.

TABLE I: Amino Acid Compositions of Peptides Obtained from HA2 by Treatment with Succinic Anhydride followed by Digestion with Trypsin.[a]

	ST1	ST2	ST3a	ST3b
Hydroxyproline	4.6 (5)	1.3 (1)	6.7 (7)	8.3 (9)
Aspartic acid	3.0 (3)	1.0 (1)	2.1 (2)	2.1 (2)
Threonine	1.0 (1)		1.9 (2)	1.9 (2)
Serine	2.9 (3)	0.2	4.0 (4)	3.9 (4)
Glutamic acid	6.1 (6)	1.3 (1)	3.1 (3)	3.2 (3)
Proline	7.1 (7)	2.3 (2)	5.5 (5)	7.3 (7)
Glycine	18.0 (18)	6.5 (6)	18.0 (18)	20.8 (21)
Alanine	5.0 (5)	3.0 (3)	5.3 (5)	7.2 (7)
Valine	1.1 (1)	1.0 (1)		
Leucine		0.2	2.0 (2)	1.9 (2)
Phenylalanine		0.9 (1)		
Hydroxylysine	0.1		0.2	0.2
Lysine	3.1[b] (4)	0.8[b] (1)	2.5[b] (3)	2.5[b] (3)
Arginine	1.0 (1)	1.1 (1)	2.1 (2)	2.9 (3)
Total residues	53.0 (54)	19.6 (18)	54.3 (53)	62.2 (63)

[a] Values are expressed as residues per peptide. A space indicates 0.1 residue or less. Numbers in parentheses indicate theoretical compositions based on the extent of overlap indicated in Figure 1. [b] Low values are due to incomplete release of lysine after acid hydrolysis of the succinylated peptide.

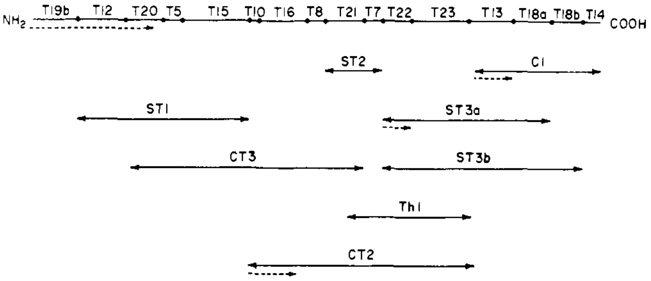

FIGURE 1: The order of the tryptic peptides in HA2. The solid lines represent sequences identified by amino acid composition and the broken lines represent the extent of Edman degradation performed with the automatic sequencer. ST1, ST2, ST3a, and ST3b were obtained by succinylation and subsequent tryptic digestion of HA2. Th1 and C1 represent thermolysin-and chymotrypsin-produced sequences, respectively. CT2 and CT3 are peptides obtained by combined chymotryptic and tryptic digestion.

sequence of the remaining tryptic peptides, except for T19b, T13, and the partial sequence of T12, was obtained by subtractive Edman degradation using a procedure described previously (Balian *et al.*, 1971). The internal sequence of peptides T19b, T13, and T12 was obtained using a Beckman Model 890A automatic sequencer with a modification of the procedure of Edman and Begg (1967) (Hermodson *et al.*, 1972).

Succinylation. HA2 (30 mg) was dissolved in 5 ml of 1 M Tris buffer at pH 11.0. The solution was warmed to 37° for 30 min then placed in an ice bath at 4°. Succinic anhydride (320 mg) was added slowly over a period of 1 hr and the pH was maintained at 8.0 with 5 N NaOH. The reaction was allowed to proceed for a total of 2 hr. The succinylated peptide was dialyzed against several changes of 0.005 N NH_4OH, pH 10, at 4°, lyophilized, and digested with trypsin. The resulting peptides were separated by molecular sieve chromatography on Sephadex G-50 (Pharmacia Fine Chemicals) in 0.2 M NH_4HCO_3 (pH 7.8) and by ion-exchange column chromatography.

Amino acid analysis was performed on a Beckman 120C analyzer using a single-column gradient elution system (Miller and Piez, 1966). A range card was used on the recorder for high sensitivity. Corrections for hydrolytic losses and incomplete release of valine have been reported (Bornstein, 1970).

Results

Order of Tryptic Peptides in HA2. Cleavage of α1-CB8-HA2 with trypsin produced 16 unique peptides (Figure 1). The order of the first three peptides was established by determination of the amino acid sequence from the NH_2 terminus of HA2 using the automatic sequencer. The following sequence was obtained: Gly-Ala-Hyp-Gly-Ile-Ala-Gly-Ala-Hyp-Gly-Phe-Hyp-Gly-Ala-Arg-Gly-Pro-Ser-Gly-Pro-Y-Gly-Pro-Y-Gly-Ala-Hyp-Gly-Pro-Lys-Gly-Asn-Y-Gly-X-Hyp-Gly-Ala-Hyp.

Since HA2 contains a single residue of isoleucine, found in T19b, peptide T19b must be NH_2 terminal in HA2. The remainder of the sequence shown above corresponds to the internal sequences of T12 and T20 (see below). The order of tryptic peptides at the NH_2 terminus of HA2 is therefore T19b-T12-T20.

Succinylation of HA2, to block ε-amino groups, followed by digestion with trypsin resulted in the isolation of four peptides designated by ST (Table I). These peptides were separated by chromatography on Sephadex G-50 and on cation- and anion-exchange resins.

The amino acid composition of ST1 indicated that it contains the peptides T12, T20, T5, and T15. Since only T15 contains arginine, it must be COOH-terminal in ST1. T12 was shown to precede T20 by sequencer data; the order of tryptic peptides in ST1 must therefore by T12-T20-T5-T15. The composition of the succinylated peptide ST2 (Table I) corresponds to that of peptides T21 plus T7. Since T7 is the arginine-containing peptide, the order of the two tryptic peptides must be T21-T7.

Digestion with chymotrypsin provided the overlap necessary to order the tryptic peptides at the COOH terminus of HA2. The composition of Cl (Table II) corresponds to that of the sum of peptides T13 (less Gly-Leu), T18a, T18b, and T14. Since T14 contains homoserine, it must be COOH terminal in α1-CB8 and HA2. Using the automatic sequencer the first 12 residues at the NH_2 terminus of Cl were found to be Thr-Gly-Ser-Hyp-Gly-Ser-Hyp-Gly-Pro-Asp-Gly-Lys. This corresponds to the greater part of the sequence of T13, showing that the order of the tryptic peptides is Cl is T13-(T18a, T18b)-T14.

The amino acid composition of the peptide ST3a (Table I) resulting from tryptic digestion of succinylated HA2 corresponds to that of the sum of tryptic peptides T22, T23, T13, and T18a. In addition the peptide ST3b, corresponding to the tryptic peptides T22, T23, T13, T18a, and T18b, was isolated. These two overlapping peptides result from incomplete tryptic cleavage of the Arg–Hyp bond in T18 and indicate that T18a precedes T18b in HA2. Partial cleavage with trypsin of Arg–Hyp bonds in collagen sequences was previously reported by Highberger *et al.* (1971).

The NH_2-terminal sequence of ST3a was determined using

TABLE II: Amino Acid Compositions of Peptides Produced by Digestion of HA2 with Chymotrypsin and Thermolysin.[a]

	$C1^b$	CT2	CT3	Thl
Hydroxyproline	4.2 (5)	7.9 (9)	7.4 (9)	4.1 (4)
Aspartic acid	2.0 (2)	1.1 (1)	3.1 (3)	0.2
Threonine	1.9 (2)		1.2 (1)	
Serine	2.1 (2)	3.6 (4)	3.4 (3)	2.1 (2)
Homoserine	0.5^b (1)			
Glutamic acid	2.5 (2)	5.2 (4)	7.2 (7)	3.0 (3)
Proline	5.5 (5)	6.5 (6)	6.4 (6)	4.0 (4)
Glycine	13.5 (13)	26.3 (25)	24.3 (25)	13.0 (13)
Alanine	4.2 (4)	7.7 (8)	8.0 (7)	6.0 (6)
Valine	0.7^b (1)	1.0 (1)	1.8 (2)	1.0 (1)
Leucine		2.9 (3)	1.1 (1)	1.0 (1)
Phenylalanine		0.9 (1)	0.9 (1)	
Hydroxylysine		0.1		0.1
Lysine	1.0 (1)	2.9 (3)	3.6 (4)	2.4 (3)
Arginine	2.0 (2)	4.9 (5)	4.1 (4)	2.1 (2)
Total residues	40.1 (40)	71.0 (70)	72.5 (73)	39.0 (39)

[a] Values are expressed as residues per peptide. A space indicates 0.1 residue or less. Numbers in parentheses indicate theoretical compositions based on the extent of overlap indicated in Figure 1. Peptides CT2 and CT3 result from the combined action of chymotrypsin and trypsin. [b] Low values of Hse and Val result from partial cleavage by chymotrypsin at the Gln–Ala bond in T14 with loss of the sequence Ala-Gly-Val-Hse.

TABLE III: Amino Acid Composition of T12 and Its Collagenase-Produced Peptides.[a]

		Collagenase Peptides			
	T12	A	B^b	C	D
Hydroxyproline	1.0	0.8	(0.4)		
Serine	2.0	0.9		1.0	
Glutamic acid	1.1		1.0	(0.3)	
Proline	3.9	1.0	1.0	1.2	1.0
Glycine	5.2	2.1	1.4	1.3	1.0
Alanine	1.2	1.0	(0.5)		
Lysine	1.0				1.1
Hydroxylysine	0.1				
Total residues	15	6	3	3	3

[a] Values are expressed as residues per peptide. A space indicates 0.1 residue or less. Residues in parentheses are fractional residues thought to be impurities. [b] Partial cleavage at the Ser-Gly bond in peptide A (below) results in the formation of the tripeptide (Gly,Ala,Hyp) which was incompletely resolved from peptide B and accounts for the presence of Hyp and Ala in this analysis.

Gly-Pro-Ser↓Gly-Pro-Gln↓Gly-Pro-Ser↓Gly-Ala-Hyp↓Gly-Pro-Lys
←C→←B→←A→←D→

the automatic sequencer and found to be Gly-Ser-Hyp-Gly-Pro-Ala-Gly-Pro-X-Gly, a sequence which corresponds to the internal sequence of T22. The order of tryptic peptides in ST3b must therefore be T22-T23-T13-T18a-T18b. Digestion of HA2 with thermolysin produced peptide Thl (Table II) with a composition corresponding to that of the COOH-terminal pentapeptide from T21 (Val-Ala-Gly-Pro-Lys) plus T7, T22, T23 and Gly from T13 (see Internal Sequences of the Tryptic Peptides). The order of the tryptic peptides in Thl must therefore be Val-Ala-Gly-Pro-Lys(T21)-T7-T22-T23-Gly(T13).

The chymotrypsin-produced peptide CT2 provided the overlap necessary to order the three remaining arginine-containing peptides T10, T16, and T8. The amino acid composition of CT2 (Table II) indicates that it includes (T10, T16, T8)-T21-T7-T22-T23-Gly-Leu(T13). This peptide resulted from cleavage of the Arg–Gly bond linking T15 and T10 (presumably due to a small amount of tryptic activity in the purified chymotrypsin preparation) and a chymotryptic split of the Leu–Thr bond in T13. The NH_2-terminal sequence of CT2 was determined using the automatic sequencer and was found to be Gly-Ala-Arg-Gly-Glu-Hyp-Gly-Pro-Ser-Gly-Leu-Hyp-Gly-Pro-Hyp. This sequence orders the initial tryptic peptides in CT2, T10 and T16, and restricts T8 to a position between T16 and T21. The order of tryptic peptides in CT2 is therefore T10-T16-T8-T21-T7-T22-T23-Gly-Leu (T13).

An additional peptide, CT3, was isolated from a chymotrypsin digest of HA2 and had a composition consistent with the sum of T20 (less Gly-Asn), T5, T15, T10, T16, T8, and T21 (Table II). The order of the tryptic peptides in HA2 is therefore established as T19b-T12-T20-T5-T15-T10-T16-T8-T21-T7-T22-T23-T13-T18a-T18b-T14 (Figure 1).

Internal Sequences of the Tryptic Peptides

T19: Gly-Ala-Hyp-Gly-Ile-Ala-Gly-Ala-Hyp↓Gly-Phe-Hyp-Gly-Ala-Arg (Residues 1–15).[2] The sequence of this peptide was obtained by Edman degradation of HA2 using the automatic sequencer. Digestion of T19b with collagenase yielded two major peptides which were separated by paper electrophoresis at pH 3.6. The amino acid composition of the two collagenase peptides was found to be (Gly_3,Ala_3,Hyp_2,Ile) and (Gly_2,Phe,Hyp,Ala,Arg). A thermolysin digest of α1-CB8 yielded an isoleucine-containing peptide with the composition (Ile,Ala_2,Gly_3,Hyp_2,Phe). These internal peptides from T19b are consistent with the sequence of the tryptic peptide.

T12: Gly-Pro-Ser↓Gly-Pro-Gln↓Gly-Pro-Ser↓Gly-Ala-Hyp↓Gly-Pro-Lys (Residues 16–30). The partial sequence of this peptide was determined by degradation of HA2 with the automatic sequencer and was found to be: Gly-Pro-Ser-Gly-Pro-Y-Gly-Pro-Y-Gly-Ala-Hyp-Gly-Pro-Lys. Digestion of T12 with collagenase yielded four major peptides which were separated by cation-exchange chromatography. Amino acid analyses (Table III) indicated that peptide C is the NH_2-terminal peptide; the lysine-containing peptide, D, must be COOH terminal. The sequencer results shown above indicate that collagenase peptide A, which contains one residue of alanine, must follow peptide B, thus confirming the position of Glx as residue 21 and Ser as residue 24. Collagenase B was

[2] The arrow in this and subsequent headings indicates a point of cleavage by collagenase.

TABLE IV: Amino Acid Composition of T20 and Its Papain-Produced Peptides.[a]

		Papain Peptides		
	T20	D	F	H
Hydroxyproline	1.9	1.0	0.8	
Aspartic acid	2.0	1.0	1.3	1.0
Serine	0.8	1.0		
Glutamic acid	0.9	1.0		
Glycine	4.5	3.2	1.3	
Alanine	1.3		1.0	
Lysine	1.0		0.8	1.0
Total residues	12	7	5	2

[a] Values are expressed as residues per peptide.

found to be neutral at pH 6.5 indicating that residue 21 is glutamine.

T20: Gly-Asn-Ser-Gly-Glu-Hyp-Gly-Ala-Hyp-Gly-Asn-Lys (Residues 31–42). The NH_2-terminal sequence of T20, obtained by subtractive Edman degradation, was found to be Gly-Asx-Ser. Digestion of T20 with papain yielded three major peptides (Table IV) which were separated by cation-exchange chromatography. The sequence of peptide D was found to be Gly-Asx-Ser-Gly-Glx-Hyp-Gly. This corresponds to the NH_2-terminal sequence of T20. Papain peptides F and H contain lysine and must be overlapping peptides derived from the COOH terminus of T20 (Figure 2). Asparagine was assigned to position 32 by identification of the PTH-amino acid by gas–liquid chromatography. Papain D was acidic at pH 6.5 after two subtractive Edman degradations showing that residue 35 is glutamic acid. Asparagine was assigned to position 41 since papain H was basic at pH 6.5.

T5: Gly-Asp-Thr-Gly-Ala-Lys (Residues 43–48). The sequence of this hexapeptide was obtained by subtractive Edman degradation and by the dansyl-Edman procedure. The intact peptide was neutral at pH 6.5 showing that residue 44 is aspartic acid.

T15: Gly-Glu-Hyp-Gly-Pro-Ala-Gly-Val-Gln↓-Gly-Pro-Hyp↓-Gly-Pro-Ala-Gly-Glu-Glu-Gly-Lys-Arg (Residues 49–69).*[3] The NH_2-terminal sequence of the intact peptide was found to be Gly-Glx-Hyp* by subtractive Edman degradation. Another tryptic peptide (T17) had the same amino acid composition as T15 less arginine and had the NH_2-terminal sequence Gly-Glx-Hyp*-Gly-Pro-Ala-Gly-Val-Glx-Gly. Thus the arginyl residue is located at the carboxyl end of T15. Both T15 and T17 were isolated from a tryptic digest of HA2 due to incomplete tryptic cleavage of the Lys–Arg bond in T15. Arginine was also isolated as T11 (Bornstein, 1970). Digests of T15 and T17 with collagenase yielded identical peptides except for peptides derived from the COOH-terminal sequences. On this basis and on the basis of yields of T15, T17, and T11 relative to the other tryptic peptides, it is concluded that T15 and T17 are derived from the same sequence in HA2.

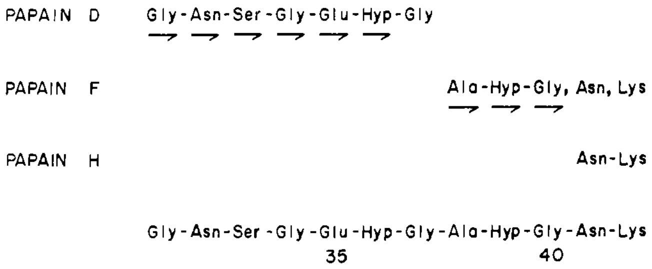
PAPAIN D
Gly-Asn-Ser-Gly-Glu-Hyp-Gly
PAPAIN F
Ala-Hyp-Gly, Asn, Lys
PAPAIN H
Asn-Lys
Gly-Asn-Ser-Gly-Glu-Hyp-Gly-Ala-Hyp-Gly-Asn-Lys
35
40

FIGURE 2: Amino acid sequence of T20. Horizontal arrows indicate the extent of Edman degradation.

The internal sequence of T15 was determined as follows (see Figure 3). The NH_2-terminal sequences of collagenases A and B were identical and correspond to the NH_2 terminus of T15. Collagenase A contains an additional sequence Gly-Pro-Hyp (collagenase C) which must follow collagenase B. Collagenase D contained lysine and arginine and must be derived from the COOH-terminal sequence in T15. Subtractive Edman degradation of collagenase D revealed the sequence Gly-Pro-Ala-Gly,Glx,Glx,Gly,Lys,Arg. Papain digestion of the COOH-terminal collagenase peptide from T17 yielded two peptides E1 and E2. The remaining sequence in T15 was obtained by subtractive Edman degradation of papain E2.

The amide assignment was made as follows. Residues 50 and 57 were identified as glutamic acid and glutamine, respectively, since collagenase B was acidic at pH 6.5 and became neutral after two Edman degradations. Glutamic acid was assigned to residues 65 and 66 since collagenase D from T17 was acidic at pH 6.5.

T10: Gly-Ala-Arg (Residues 70–72). The sequence of this tripeptide was obtained by the dansyl-Edman procedure.

T16: Gly-Glu-Hyp-Gly-Pro-Ser↓-Gly-Leu-Hyp↓-Gly-Pro-Hyp-Gly-Glu-Arg (Residues 73–87). Subtractive Edman degradation of this peptide gave the sequence Gly-Glx-Hyp-Gly-Pro-Ser-Gly-Leu-Hyp-Gly. Digestion of T16 with collagenase yielded three major peptides (Table V) which were separated by cation-exchange chromatography. The amino acid composition of peptide A corresponds to the first nine residues of T16. The composition of collagenase B is the same as A but lacks the triplet Gly-Leu-Hyp. Subtractive Edman degradation of collagenase C gave the COOH-terminal sequence of

TABLE V: Amino Acid Composition of T16 and Its Collagenase-Produced Peptides.[a]

		Collagenase Peptides		
	T16	A	B	C
Hydroxyproline	3.0	2.1	1.2	0.9
Serine	1.1	1.1	1.0	
Glutamic acid	2.1	0.9	1.1	1.1
Proline	1.9	0.9	1.0	1.0
Glycine	4.9	3.1	2.1	2.1
Leucine	1.0	0.9		
Arginine	0.9			1.0
Total residues	15	9	6	6

[a] Values are expressed as residues per peptide.

[3] The symbol Hyp* is used to indicate an incompletely hydroxylated prolyl residue.

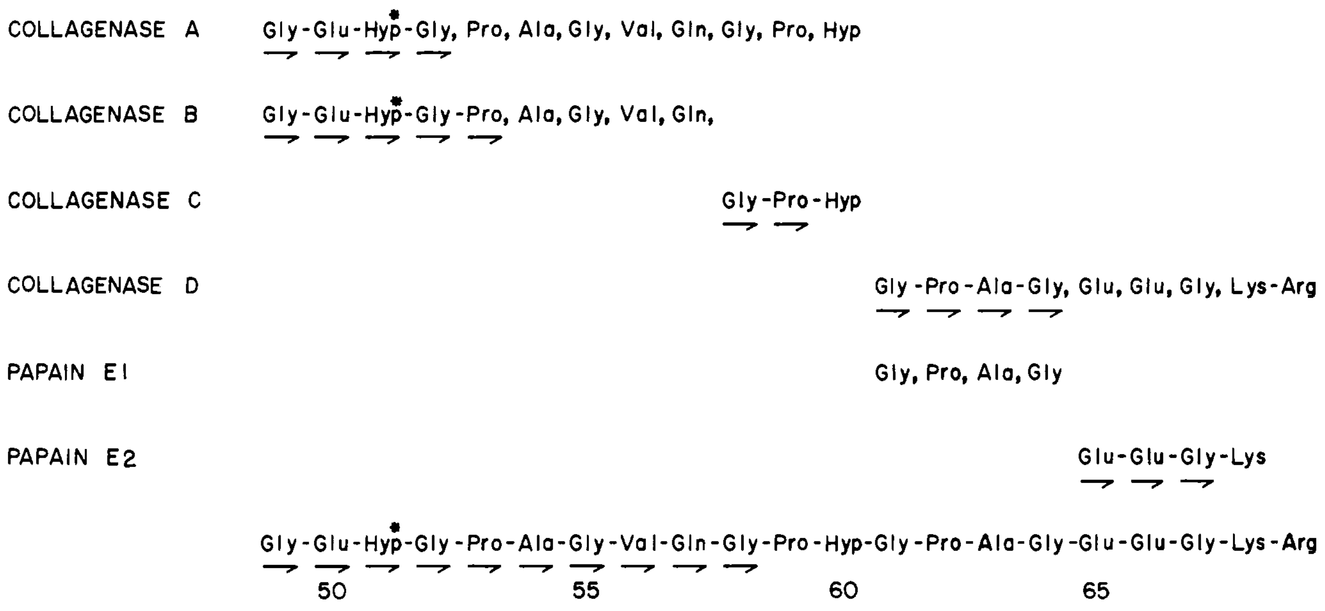
COLLAGENASE A
Gly-Glu-Hyp*-Gly, Pro, Ala, Gly, Val, Gln, Gly, Pro, Hyp
COLLAGENASE B
Gly-Glu-Hyp*-Gly-Pro, Ala, Gly, Val, Gln,
COLLAGENASE C
Gly-Pro-Hyp
COLLAGENASE D
Gly-Pro-Ala-Gly, Glu, Glu, Gly, Lys-Arg
PAPAIN E1
Gly, Pro, Ala, Gly
PAPAIN E2
Glu-Glu-Gly-Lys
Gly-Glu-Hyp*-Gly-Pro-Ala-Gly-Val-Gln-Gly-Pro-Hyp-Gly-Pro-Ala-Gly-Glu-Glu-Gly-Lys-Arg
50
55
60
65

FIGURE 3: Amino acid sequence of T15. Collagenase- and papain-produced peptides are arranged in order and show the extent of sequence overlap obtained. Papain E1 and E2 were obtained from T17, which lacks arginine. The horizontal arrows indicate the extent of Edman degradation. The prolyl residue in position 51 is incompletely hydroxylated (Hyp*).

TABLE VI: Amino Acid Composition of T23 and Its Collagenase- and Papain-Produced Peptides.[a]

		Collagenase Peptides		Papain Peptides						
	T23	II	IV	A	B	C	E	F	G	H
Hydroxyproline	2.5	2.9		0.8	0.9	1.0		0.8		0.9
Serine	1.1	1.0		1.1	1.0					
Glutamic acid	2.2	2.1		1.1			1.0	1.4		
Proline				0.3				0.3		0.1
Glycine	6.5	5.3	1.1	3.0	2.2	1.2	1.0	2.5		1.2
Alanine	3.2	2.2	1.0	0.9			1.0	1.0	1.0	
Leucine	1.1	1.0				0.9				
Lysine	1.0		0.9						1.0	
Arginine	0.9	1.0						1.0		0.9
Total residues	18	15	3	7	4	3	3	6	2	3

[a] Values are expressed as residues per peptide.

T16. Glutamic acid was assigned to residues 74 and 86 since collagenase A was acidic and collagenase C was neutral at pH 6.5.

T8: Gly-Gly-Hyp-Gly-Ser-Arg (*Residues 88–93*). The amino acid composition of this hexapeptide was Gly_3,Hyp,Ser,Arg (Bornstein, 1970) indicating that it must contain a Gly-Gly sequence. The sequence of T8 was determined by subtractive Edman degradation.

T21: Gly-Phe-Hyp-Gly-Ala-Asp-Gly-Val-Ala-↓Gly-Pro-Lys (*Residues 94–105*). Subtractive Edman degradation of this peptide gave the amino-terminal sequence Gly-Phe-Hyp-Gly-Ala-Asx. Digestion of T21 with collagenase produced two peptides, an NH_2-terminal nonapeptide (collagenase A) and a COOH-terminal tripeptide (collagenase B) which were resolved by ion-exchange chromatography. Collagenase A had a composition Gly_3,Phe,Hyp,Ala_2,Asp,Val; its COOH-terminal sequence was determined as Val-Ala by digestion with carboxypeptidase A. Edman degradation of collagenase B gave the sequence Gly-Pro-Lys. Collagenase A was acidic at pH 6.5 indicating that residue 99 is aspartic acid.

T7: Gly-Pro-Ala-Gly-Glu-Arg (*Residues 106–111*). The sequence of this peptide was obtained by subtractive Edman degradation. Glutamic acid was assigned to position 110 since the intact peptide was neutral at pH 6.5.

T22: Gly-Ser-Hyp-↓Gly-Pro-Ala-↓Gly-Pro-Lys (*Residues 112–120*). The sequence of this tryptic peptide was obtained by degradation of peptide ST3a with the automatic Sequencer (Figure 1). The NH_2-terminal sequence of the isolated tryptic peptide was found to be Gly-Ser-Hyp. In addition, digestion with collagenase yielded three peptides which were separated by ion-exchange chromatography and had compositions consistent with the above sequence.

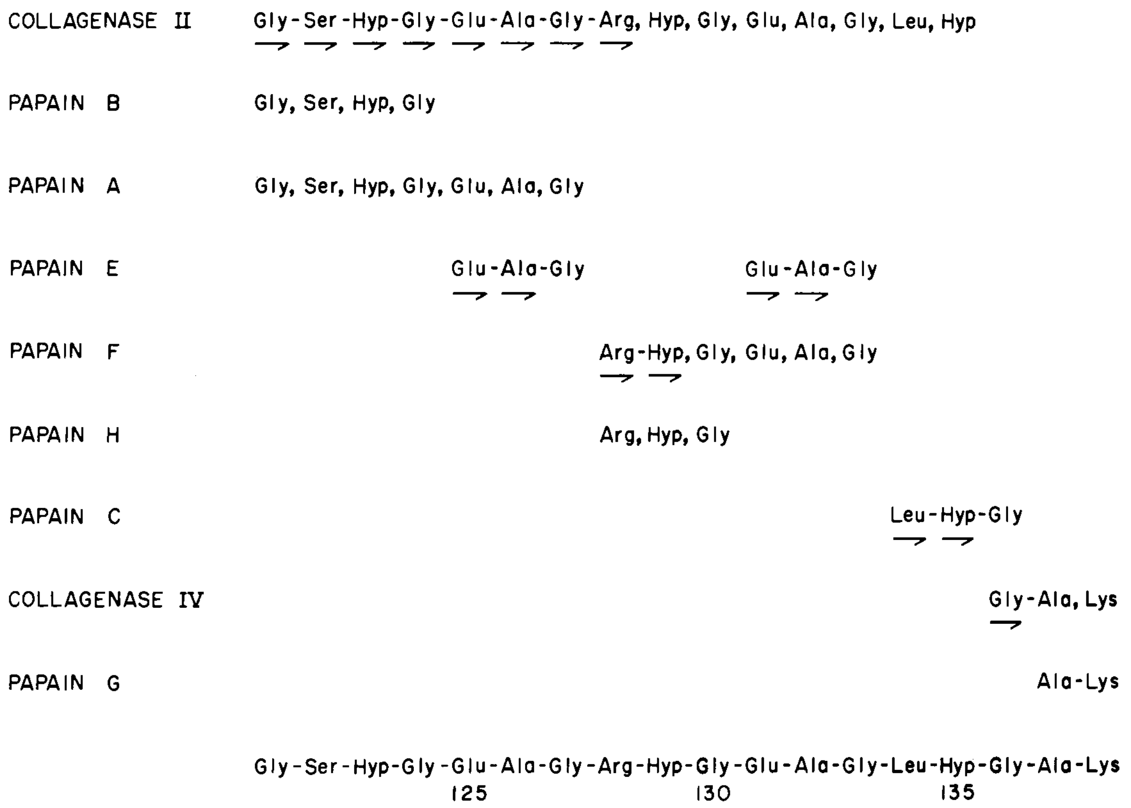
COLLAGENASE II
Gly-Ser-Hyp-Gly-Glu-Ala-Gly-Arg, Hyp, Gly, Glu, Ala, Gly, Leu, Hyp
PAPAIN B
Gly, Ser, Hyp, Gly
PAPAIN A
Gly, Ser, Hyp, Gly, Glu, Ala, Gly
PAPAIN E
Glu-Ala-Gly
Glu-Ala-Gly
PAPAIN F
Arg-Hyp, Gly, Glu, Ala, Gly
PAPAIN H
Arg, Hyp, Gly
PAPAIN C
Leu-Hyp-Gly
COLLAGENASE IV
Gly-Ala, Lys
PAPAIN G
Ala-Lys
Gly-Ser-Hyp-Gly-Glu-Ala-Gly-Arg-Hyp-Gly-Glu-Ala-Gly-Leu-Hyp-Gly-Ala-Lys
125
130
135

FIGURE 4: Amino acid sequence of T23. Collagenase- and papain-produced peptides are arranged in order and indicate the extent of sequence overlap obtained. Horizontal arrows indicate the extent of Edman degradation.

T23: Gly-Ser-Hyp-Gly-Glu-Ala-Gly-Arg-Hyp-Gly-Glu-Ala-Gly-Leu-Hyp↓-Gly-Ala-Lys (*Residues 121–138*). Subtractive Edman degradation of the intact peptide revealed the NH_2-terminal sequence, Gly-Ser. Cleavage with collagenase produced two peptides (Table VI) which were separated both by electrophoresis at pH 6.5 and by ion-exchange chromatography. Collagenase II, which contains one residue of serine, must be derived from the NH_2 terminus of T23 (Figure 4) and collagenase IV (Gly-Ala-Lys) must be COOH terminal. Edman degradation of collagenase II revealed the sequence Gly-Ser-Hyp-Gly-Glu-Ala-Gly-Arg, Hyp_2Gly_2,Glu,Ala,Leu.

Digestion of T23 with papain yielded seven peptides which were separated by ion-exchange chromatography (Table VI). The amino acid composition of papain A indicated that it originates from the NH_2 terminus of T23 and overlaps papain peptides B (Gly,Ser,Hyp,Gly) and E (Glx-Ala-Gly). Since T23 contains two residues of glutamic acid, peptide E must originate from repeating sequences in T23 (Figure 4). Papain F overlaps peptides E and H. Papain F has an NH_2-terminal sequence Arg-Hyp and must follow papain A in T23, thus restricting papain C to the position between papain F and collagenase IV (Figure 4). Glutamic acid was assigned to residues 125 and 131 due to the negative charge, at pH 6.5, of collagenase peptide II.

T13: Gly-Leu-Thr-Gly-Ser-Hyp-Gly-Ser-Hyp-Gly-Pro-Asp-Gly-Lys (*Residues 139–152*). Subtractive Edman degradation revealed the NH_2-terminal sequence of T13 to be Gly-Leu-Thr. Cleavage of HA2 with chymotrypsin resulted in a split at the Leu-Thr bond in T13. The resulting peptide, Cl, was analyzed by the automatic sequencer (see Order of Tryptic Peptides and Figure 1). This analysis provided the remaining sequence of T13. Aspartic acid was assigned to position 150 by identification of the PTH derivative.

T18a: Thr-Gly-Pro-Hyp-Gly-Pro-Ala-Gly-Glx-Asx-Gly-Arg (*Residues 153–164*). Ten consecutive subtractive Edman degradations in this peptide established the sequence of the first ten residues. Digestion with carboxypeptidases B and A revealed the COOH-terminal sequence to be Gly-Arg thus completing the sequence of this peptide.

The intact peptide was neutral at pH 6.5 indicating that one of the two acidic amino acids existed as the amide. However, the tripeptide Glx-Asx-Gly, obtained by papain digestion of the overlapping peptide Cl, was not retarded by cation-exchange columns and attempts at Edman degradation of this tripeptide were unsuccessful, suggesting that cyclization to pyrrolidone-5-carboxylic acid had occurred following proteolytic digestion. Since glutaminyl peptides cyclize far more readily than glutamyl peptides, a tentative assignment of Gln to position 161 and Asp to position 162 can be made. Such an assignment would also be consistent with the lack of appreciable cleavage by hydroxylamine at the Asp–Gly bond (residues 162–163). However, since numerous attempts to substantiate these assignments produced inconclusive results, positions 161 and 162 were designated as Glx and Asx, respectively.

T18b: Hyp-Gly-Pro-Ala↓-Gly-Pro-Hyp↓-Gly-Ala-Arg* (*Residues 165–174*). Nine subtractive Edman degradations yielded the sequence shown above. Moreover, digestion with collagenase produced a tetrapeptide (Hyp,Gly,Pro,Ala) and two tripeptides (Gly,Pro,Hyp) and (Gly,Ala,Arg) which provided additional support for the above sequence.

T14: Gly-Gln-Ala-Gly-Val-Hse (*Residues 175–180*). The sequence of T14 was obtained by subtractive Edman degra-

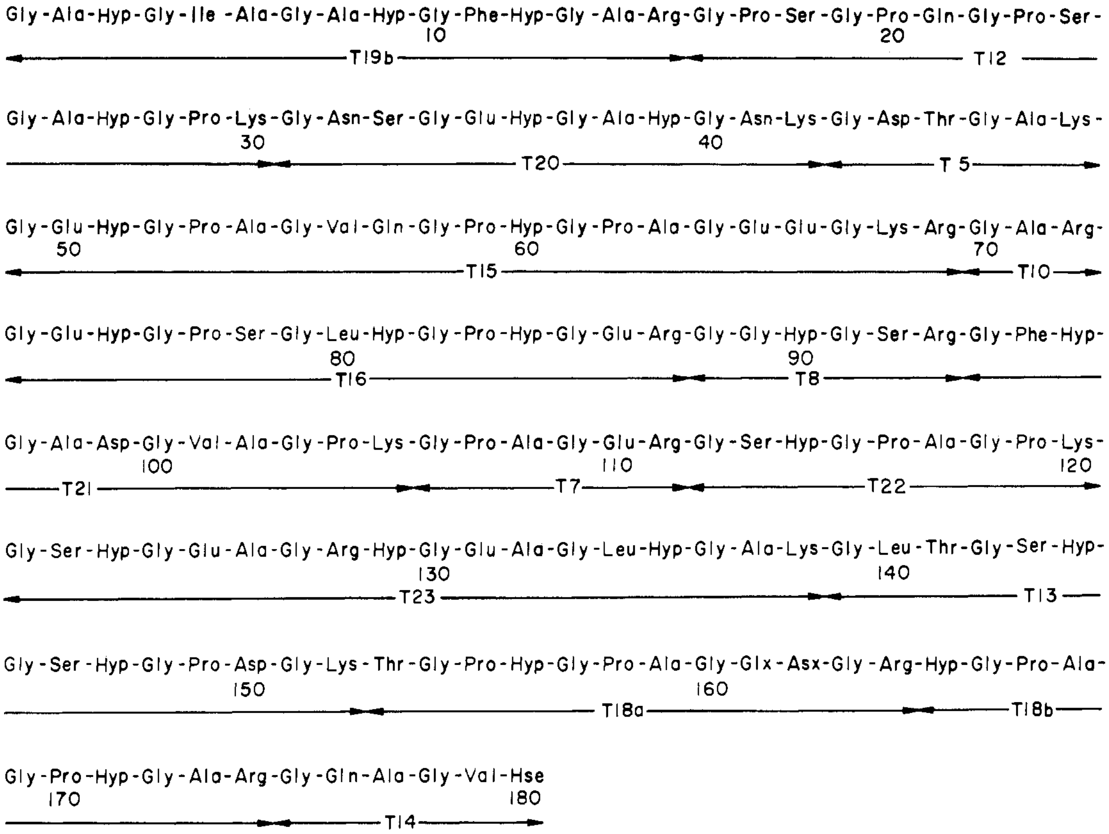
Gly-Ala-Hyp-Gly-Ile-Ala-Gly-Ala-Hyp-Gly-Phe-Hyp-Gly-Ala-Arg-Gly-Pro-Ser-Gly-Pro-Gln-Gly-Pro-Ser-
10
20
T19b
T12
Gly-Ala-Hyp-Gly-Pro-Lys-Gly-Asn-Ser-Gly-Glu-Hyp-Gly-Ala-Hyp-Gly-Asn-Lys-Gly-Asp-Thr-Gly-Ala-Lys-
30
40
T20
T 5
Gly-Glu-Hyp-Gly-Pro-Ala-Gly-Val-Gln-Gly-Pro-Hyp-Gly-Pro-Ala-Gly-Glu-Glu-Gly-Lys-Arg-Gly-Ala-Arg-
50
60
70
T15
T10
Gly-Glu-Hyp-Gly-Pro-Ser-Gly-Leu-Hyp-Gly-Pro-Hyp-Gly-Glu-Arg-Gly-Gly-Hyp-Gly-Ser-Arg-Gly-Phe-Hyp-
80
90
T16
T8
Gly-Ala-Asp-Gly-Val-Ala-Gly-Pro-Lys-Gly-Pro-Ala-Gly-Glu-Arg-Gly-Ser-Hyp-Gly-Pro-Ala-Gly-Pro-Lys-
100
110
120
T21
T7
T22
Gly-Ser-Hyp-Gly-Glu-Ala-Gly-Arg-Hyp-Gly-Glu-Ala-Gly-Leu-Hyp-Gly-Ala-Lys-Gly-Leu-Thr-Gly-Ser-Hyp-
130
140
T23
T13
Gly-Ser-Hyp-Gly-Pro-Asp-Gly-Lys-Thr-Gly-Pro-Hyp-Gly-Pro-Ala-Gly-Glx-Asx-Gly-Arg-Hyp-Gly-Pro-Ala-
150
160
T18a
T18b
Gly-Pro-Hyp-Gly-Ala-Arg-Gly-Gln-Ala-Gly-Val-Hse
170
180
T14

FIGURE 5: The complete amino acid sequence of α1-CB8-HA2. Only the tryptic peptides are indicated. Prolyl residues at positions 51 and 165 are incompletely hydroxylated as indicated by Hyp*.

dation. Glutamine was assigned to residue 176 since the peptide was basic at pH 6.5. Absence of a negative charge presumably results from lactone formation by the COOH-terminal homoseryl residue.

The complete amino acid sequence of HA2 is summarized in Figure 5. Table VII summarizes the basis for the amide assignments in HA2. In agreement with other collagen sequences from the helical region of collagen (Bornstein, 1967; Butler, 1970; Highberger *et al.*, 1971; Butler and Ponds, 1971; von der Mark *et al.*, 1970; Balian *et al.*, 1971), glycine is present as every third residue and hydroxyproline is restricted to position Y in the repeating sequence Gly-X-Y. In addition, hydroxylation of prolyl residues at residues 51 and 165 is incomplete. The distribution of other amino acids is considered in the Discussion.

Discussion

The determination of the primary structure of α1-CB8-HA2 extends the known sequence of the α1 chain of rat collagen (Kang *et al.*, 1967; Bornstein, 1967; Butler, 1970; Butler and Ponds, 1971; Balian *et al.*, 1971) to 418 residues from the NH_2 terminus. This sequence together with the sequence of α1-CB6 from bovine collagen (Wendt *et al.*, 1972; Fietzek *et al.*, 1972) and rat α2-CB2 (Highberger *et al.*, 1971) provides a basis for the examination of the distribution of amino acids in the repetitive collagen structure (Table VIII). The previously observed restriction of leucine and phenylalanine to position X in the triplet sequence Gly-X-Y (Balian *et al.*, 1971) is confirmed and extended by these data. The nonrandom distribution of leucine is particularly striking since the probability, on the basis of chance, of 13 consecutively determined leucines occurring in position X is 1.2×10^{-4}. The basis for this selection is not apparent, particularly since other hydrophobic side chains, including isoleucine and valine, occur in both positions X and Y (Table VIII). The presently accepted model for the collagen helix (Traub *et al.*, 1969) does not place restrictions on the distribution of amino acids in collagen (other than the requirement for glycine in every third position), but interactions between side chains in the triple-stranded molecule and between molecules during aggregation may dictate the observed distribution.

The relative frequency of glutamic acid in position X and threonine and arginine in position Y also seems significantly greater than that expected on the basis of chance (Table VIII). The absence of cysteine, tryptophan, and tyrosine from the known helical regions of mammalian collagens provides an additional example of the restriction in the distribution of amino acids in collagen. Presumably the properties of these amino acids do not contribute to, or may detract from, the desired functional properties of the protein. Similarly, despite the high glycine content of collagen, only one glycyl residue has been encountered in position X or Y (residue 89 in α1-CB8-HA2) in the more than 600 amino acids sequenced in the helical region of collagen. On a random basis, approximately 20 glycyl residues would have been expected in these positions. Presumably the higher degree of flexibility in the polypeptide chain which results from a lack of a substituent on the α carbon is not desirable in the collagen structure, this despite the absolute requirement for glycine in every third position in chains which form a triple helix. Furthermore, a residue such as glycine would contribute little to the inter-

TABLE VII: Assignment of Acid and Amide Groups in α1-CB8-HA2.

Residue	Assignment	Method of Determination
21	Gln	Hve,[a] T12-collagenase B
32	Asn	PTH[b]
35	Glu	Hve, T20-papain D after two Edman degradations
41	Asn	Hve, T20-papain H
44	Asp	Hve, T5
50	Glu	Hve, T15-collagenase B before and after two Edman degradations
57	Gln	
65	Glu	Hve, T17-collagenase D
66	Glu	
74	Glu	Hve, T16-collagenase A
86	Glu	Hve, T16-collagenase C
99	Asp	Hve, T21-collagenase A
110	Glu	Hve, T7
125	Glu	Hve, T23-collagenase II
131	Glu	
150	Asp	PTH
161	Glx	
162	Asx	
166	Gln	Hve, T14

[a] High-voltage electrophoresis at pH 6.5. [b] Identification of the PTH derivative of the amino acid by gas–liquid chromatography.

TABLE VIII: Distribution of Amino Acids in the Collagen Triplet Gly-X-Y.[a]

	X	Y	Probability
Aspartic acid	7	7	
Asparagine	5	4	
Threonine	2	10	1.6×10^{-2}
Serine	14	12	
Glutamic acid	20	5	9.5×10^{-3}
Glutamine	4	9	
Alanine	27	35	
Valine	5	3	
Methionine[b]	1	4	
Isoleucine	2	2	
Leucine	13	0	1.2×10^{-4}
Phenylalanine	6	0	1.6×10^{-2}
Lysine[c]	6	14	
Histidine	2	0	
Arginine	8	27	1.4×10^{-3}

[a] Data compiled from Bornstein (1967), Butler (1969), Butler and Ponds (1971), Balian *et al.* (1971), Wendt *et al.* (1972), Fietzek *et al.* (1972), Highberger *et al.* (1971), and from this publication. Proline and hydroxyproline are restricted to positions X and Y, respectively (although incomplete hydroxylation yields a small fraction of proline in position Y). Glycine is limited to the first position in the triplet with the sole exception of residue 89 in α1-CB8-HA2. [b] Identified as homoserine. [c] Includes hydroxylysine.

molecular interactions which characterize the molecule in an aggregate or fiber.

When the compositions of collagen triplets are examined, patterns emerge which suggest that not only is there a preferential distribution of amino acids in position X and Y but that the frequency with which two amino acids are associated in the same triplet may in some cases exceed that expected from the amino acid content of the protein. Thus leucine appeared in the triplet Gly-Leu-Hyp in 7 of the 13 instances in which the amino acid was determined in a collagen sequence. This represents a frequency which is almost twice that expected by chance even if leucine is restricted to position X and hydroxyproline to position Y.[4] These considerations suggest that the criteria used to judge the evolutionary origin of sequences in collagen will differ from those applied to other proteins. Further information will therefore be required before the existence of short identical sequences in the collagen α1 chain can be interpreted as evidence for duplication of genetic material (Traub and Piez, 1971; Wendt *et al.*, 1972).

Despite these reservations the high degree of similarity in the amino acid compositions of the α1 and α2 chains and the marked similarity in the distribution of charged residues, as shown by electron optical analyses of renatured molecules composed of only one type of chain (Tkocz and Kühn, 1969), strongly suggest that the two chains are homologous. In order to test this hypothesis the sequence of α2-CB2 (Highberger *et al.*, 1971) was compared by computer analysis with all possible 30 amino acid sequences in the known structure of α1 (K. A. Piez, G. Balian, E. M. Click, and P. Bornstein, submitted for publication). The region in α1 which was found to possess a degree of similarity significantly greater than any other was residues 106–135 in α1-CB8-HA2. This region corresponds precisely in distance from the NH_2 terminus of the chain to α2-CB2 and an almost identical distribution of charged residues exists in the two sequences, indicating homology of the chains. Homology of the α1 and α2 chain of mammalian collagens is consistent with preliminary evidence indicating that, in invertebrates, collagens with only a single type of α chain exist (Pikkarainen *et al.*, 1968; Nordwig and Hayduk, 1969). Collagens with three identical α chains also occur in higher animals (Miller and Matukas, 1969) and in some tissues may represent the embryonic or fetal form of the protein (Miller *et al.*, 1971).

The band pattern observed in electron micrographs of segment-long-spacing aggregates of collagen, stained with uranyl acetate and phosphotungstic acid, correlates well with the distribution of charged amino acids in the protein (von der Mark *et al.*, 1970; Balian *et al.*, 1971). These observations extend to α1-CB2-HA2 (Figure 6). In Figure 6 the position of α1-CB8 and site of cleavage by hydroxylamine were calculated assuming a length of 3000 Å for the molecule and a translation of 2.91 Å/residue. The close correspondence of acidic and basic amino acids in this region of the α1 chain with the pattern of deposition of electron-dense stain in the triple-stranded molecule attests to the similarity in the distribution of charged amino acids in the α1 and α2 chains of collagen.

Hydroxylamine was shown to cleave asparaginyl–glycyl bonds in both α1-CB3 (Butler, 1969) and α1-CB8 (Bornstein, 1969a) and a mechanism which involves the formation of the

[4] Given the sequence Gly-Leu-Y the probability that the third position contains Hyp is given by the frequency of Hyp in position Y. Of the triplets examined, only 30% contained hydroxyproline, whereas 7/13 or 54% of leucine-containing triplets were identified as Gly-Leu-Hyp.

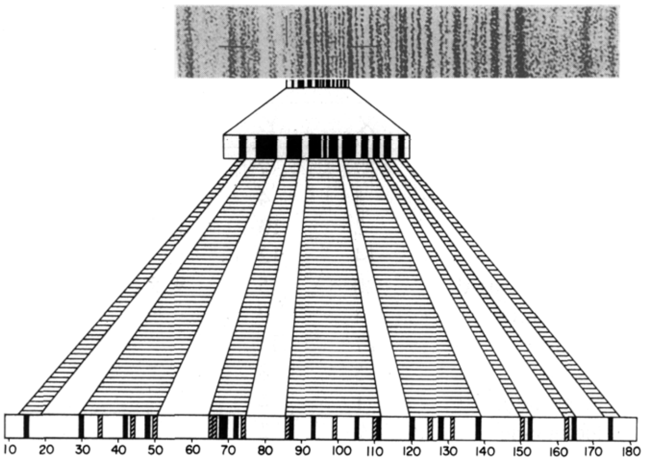

FIGURE 6: The band pattern in an electron micrograph of segment-long-spacing aggregates of calf collagen (stained with phosphotungstic acid and uranyl acetate) is compared with the distribution of acidic and basic amino acids in HA2. The NH_2 terminus of the sequence is to the left. The horizontal arrows delineate the extent of α1-CB8 and the vertical arrow indicates the point of cleavage with hydroxylamine. Hatched and solid bars represent acidic and basic amino acids, respectively. The electron micrograph was kindly provided by Dr. Klaus Kühn.

cyclic imide intermediate, anhydroaspartylglycine, has been proposed (Bornstein, 1970). The asparaginyl residues in HA2, residues 32 and 41, precede serine and lysine, respectively, making cyclic imide formation at these positions less likely. Of the four aspartyl residues in HA2, three precede glycyl residues.[5] The absence of additional hydroxylamine cleavages in significant yield in HA2 is consistent with the suggestion that amidation of the aspartyl side chain enhances the likelihood of imide formation (Bornstein, 1970). Hydroxylamine cleavage may therefore provide a useful supplement to nonenzymatic cleavage with cyanogen bromide during sequence determination of proteins.

No distinct pattern to the underhydroxylation of lysyl residues could be discerned from this work. Of the seven lysyl residues in HA2 six were found in position Y (in the triplet sequence Gly-X-Y) but none was hydroxylated to an extent exceeding 10%. Presumably the adjacent peptide sequences did not favor binding by collagen lysine hydroxylase (Kivirikko *et al.*, 1972). Similarly, the previously described incomplete hydroxylation of prolyl residues (Bornstein, 1967; Rhoads *et al.*, 1971) was observed in HA2 but a structural basis for the pattern of partial hydroxylation is not apparent.

The extensive use of bacterial collagenase during the determination of the sequence of HA1 and HA2 confirms our previous suggestion that Y–Gly bonds joining triplets containing an acidic amino acid are totally resistant to cleavage (Balian *et al.*, 1971). The enzyme preparations used in these studies undoubtedly contain a mixture of collagenases A and B (Harper and Kang, 1970). The absence of an imino acid in triplets adjacent to a potentially susceptible bond did not impart complete resistance to cleavage by collagenase but the decreased frequency of cleavage at such bonds indicates that the activity of the enzyme is markedly enhanced by the proximity of prolyl and hydroxyprolyl residues.

Acknowledgment

We are indebted to Lowell Ericcson for assistance with the automatic Sequencer, to Dr. K. Kühn for permission to use the electron micrograph in Figure 6 and for providing us with the sequence of α1-CB6 prior to publication, to Drs. Kenneth Walsh and Koiti Titani for helpful discussions and to Felicia Arguelles for skillful technical assistance.

References

Balian, G., Click, E. M., and Bornstein, P. (1971), *Biochemistry 10*, 4470.

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

Bornstein, P. (1969a), *Biochem. Biophys. Res. Commun. 36*, 957.

Bornstein, P. (1969b), *Biochemistry 8*, 63.

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

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

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

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

Butler, W. T., and Ponds, S. L. (1971), *Biochemistry 10*, 2076.

Edman, P., and Begg, G. (1967), *Eur. J. Biochem. 1*, 80.

Fietzek, P. P., Rexrodt, F., Wendt, P., Stark, M. and Kühn, K. (1972), *Eur. J. Biochem.* (in press).

Harper, E., and Kang, A. H. (1970), *Biochem. Biophys. Res. Commun. 41*, 482.

Hermodson, M. A., Ericsson, L. H., Titani, K., Neurath, H., and Walsh, K. A. (1972), *Biochemistry* (in press).

Highberger, J. H., Kang, A. H., and Gross, J. (1971), *Biochemistry 10*, 610.

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

Kivirikko, K. I., Shudo, K., Sakakibara, S., and Prockop, D. J. (1972), *Biochemistry 11*, 122.

Miller, E. J., Epstein, Jr., E. H., and Piez, K. A. (1971), *Biochem. Biophys. Res. Commun. 42*, 1024.

Miller, E. J., and Matukas, V. J. (1969), *Proc. Nat. Acad. Sci. U. S. 64*, 1264.

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

Nordwig, A., and Hayduk, U. (1969), *J. Mol. Biol. 44*, 161.

Peterkofsky, B., and Diegelmann, R. (1971), *Biochemistry 10*, 988.

Piez, K. A. (1968), *Anal. Biochem. 26*, 305.

Piez, K. A., Miller, E. J., Lane, J. M., and Butler, W. T. (1969), *Biochem. Biophys. Res. Commun. 37*, 801.

Pikkarainen, J., Rantanen, J., Vastamäki, M., Lapiaho, K., and Kulonen, E. (1968), *Eur. J. Biochem. 4*, 555.

Pisano, J. J., and Bronzert, T. J. (1969), *J. Biol. Chem. 244*, 5597.

Rauterberg, J., and Kühn, K. (1968), *FEBS (Fed. Eur. Biochem. Soc.) Lett. 1*, 230.

Rhoads, R. E., Udenfriend, S., and Bornstein, P. (1971), *J. Biol. Chem. 246*, 4138.

Tkocz, C., and Kühn, K. (1969), *Eur. J. Biochem. 7*, 454.

Traub, W., Yonath, A., and Segal, D. M. (1969), *Nature (London) 221*, 914.

Traub, W., and Piez, K. A. (1971), *Advan. Protein Chem. 25*, 243.

von der Mark, K., Wendt, P., Rexrodt, F., and Kühn, K. (1970), *FEBS (Fed. Eur. Biochem. Soc.) Lett. 11*, 105.

Wendt, P., von der Mark, K., Rexrodt, F., and Kühn, K. (1972), *Eur. J. Biochem.* (in press).

[5] The identity of the side chain in position 162 was not conclusively established, but the weight of evidence indicates that it is aspartic acid not asparagine (see Results).