Hunsicker, L. G., Ruddy, S., & Austen, K. F. (1973) *J. Immunol. 110*, 128.
Hunter, W. M., & Grenwood, F. C. (1962) *Nature (London) 194*, 495.
Kerr, M. A., & Porter, R. R. (1978) *Biochem. J. 171*, 99.
Laemmli, U. K. (1970) *Nature (London) 227*, 680.
Lesavre, P. H., Hugli, T. E., Esser, A. F., & Müller-Eberhard, H. J. (1979) *J. Immunol. 123*, 529.
Livingston, D. M. (1974) *Methods Enzymol. 34*, 723.
Mancini, G., Carbowara, A. O., & Heremans, J. F. (1965) *Immunochemistry 2*, 235.
Martin, A., Lachmann, P. J., Halbwachs, L., & Hobart, M. J. (1976) *Immunochemistry 13*, 317.
Medicus, R. G., Götze, O., & Müller-Eberhard, H. J. (1976) *Scand. J. Immunol. 5*, 1049.
Mole, J. E., Bhown, A. S., & Bennett, J. C. (1977) *Biochemistry 16*, 3507.
Niemann, M. A., Nagasawa, S., Bhown, A. S., Volanakis, J., Stroud, R. M., Bennett, J. C., & Mole, J. E. (1978) *Fed. Proc., Fed. Am. Soc. Exp. Biol. 37*, 1378.
Niemann, M. A., Hollaway, W. L., & Mole, J. E. (1979) *J. High Resolut. Chromatogr. Chromatogr. Commun. 2*, 743.
Ouchterlony, O. (1958) *Prog. Allergy 5*, 1.
Platts-Mills, T. A. E., & Ishizaka, K. (1974) *J. Immunol. 113*, 348.
Reid, K. B. M. (1979) *Biochem. J. 179*, 367.
Reid, K. B. M., & Thompson, E. O. P. (1978) *Biochem. J. 173*, 863.
Sim, R. B., Porter, R. R., Reid, K. B. M., & Gigli, I. (1977) *Biochem. J. 163*, 219.
Stroud, R. M., Volanakis, J. E., Nagasawa, S., & Lint, T. F. (1979) *Immunochem. Proteins 3*, 167.
Tack, B. F., Morris, S. C., & Prahl, J. W. (1979a) *Biochemistry 18*, 1490.
Tack, B. F., Morris, S. C., & Prahl, J. W. (1979b) *Biochemistry 18*, 1497.
Vogt, W., Dames, W., Schmior, G., & Dieminger, L. (1977) *Immunochemistry 14*, 201.
Volanakis, J. E., Schrohenloher, R. E., & Stroud, R. M. (1977) *J. Immunol. 119*, 337.
Volanakis, J. E., Bhown, A. S., Bennett, J. C., & Mole, J. E. (1980) *Proc. Natl. Acad. Sci. U.S.A.* (in press).

Covalent Structure of Collagen: Amino Acid Sequence of α1(III)-CB5 from Type III Collagen of Human Liver†

Jerome M. Seyer,* Carlo Mainardi, and Andrew H. Kang

ABSTRACT: Type III collagen was prepared from human liver by limited pepsin digestion, differential salt precipitation, and carboxymethylcellulose chromatography. Ten distinct peptides were obtained by cyanogen bromide digestion. The peptide α1(III)-CB5 was further purified by carboxymethylcellulose chromatography, and its amino acid sequence was determined. Automatic Edman degradation of intact α1(III)-CB5, tryptic and thermolytic peptides, and hydroxylamine-derived fragments was used to establish the total sequence. The mammalian collagenase site contained in the α1(III)-CB5 sequence was ascertained by digestion of native type III collagen with purified rheumatoid synovial collagenase. Collagenase cleavage occurred at a single Gly–Ile bond, one triplet before the corresponding specific cleavage site of type I collagen. The present work brings the known sequence of human liver type III collagen to include α1(III)-CB3-7-6-1-8-10-2-4-5. These correspond to the homologous region of α1(I)-CB0-1-2-4-5-8-3-7 residues 11–804.

Collagen is the most abundant extracellular connective tissue protein of most vertebrates. It exists as a triple-stranded helix of three α chains, each containing over 1000 amino acid residues (Piez, 1976; Gallop et al., 1972; Traub & Piez, 1971). At least three genetically distinct, interstitial collagens occur in mammals and are referred to as types I, II, and III.

Both type I and type III collagens occur simultaneously in most connective tissues with the exception of bone which contains only type I. Type II collagen is found solely in cartilage and vitreous humor (Miller & Lunde, 1973; Chung & Miller, 1974; Epstein, 1974; Swann et al., 1972; Stuart et al., 1979). Type III collagen contains three identical α chains of 95 000 mol wt held together by disulfide linkages. Ten CNBr peptides of type III collagen of human skin and liver have been isolated and characterized (Chung et al., 1974; Seyer & Kang, 1977), and complete amino acid sequences of eight peptides, α1(III)-CB3-7-6-1-8-10-2-4, representing the first 558 residues from the NH_2 terminus have been reported (Seyer & Kang, 1977, 1978). The present report describes the amino acid sequence of α1(III)-CB5 of human liver, a 237-residue CNBr peptide which contains the collagenase cleavage site. The sequence of only one peptide from human type III collagen, α1(III)-CB9, remains to be determined.

During the preparation of this manuscript, the entire sequence of bovine skin type III collagen was reported (Fietzek et al., 1979; Dewes et al., 1979a,b; Bentz et al., 1979; Lang et al., 1979; Allman et al., 1979). The last report represents the first complete amino acid sequence analysis of an individual collagen chain.

Materials and Methods

Preparation of α1(III)-CB5. Human cirrhotic livers were obtained after autopsy, and type III collagen was prepared as previously described (Seyer et al., 1976, 1977). The CNBr peptides were obtained by digestion of purified type III collagen in 70% formic acid and separated by ion-exchange

† From the Connective Tissue Research Laboratory, Veterans Administration Medical Center, and the Departments of Medicine and Biochemistry, University of Tennessee Center for the Health Sciences, Memphis, Tennessee 38104. *Received Novermber 29, 1979.*
* Author to whom correspondence should be addressed at the Research Service (151), VA Medical Center, Memphis, TN 38104.

0006-2960/80/0419-1583$01.00/0

chromatography on CM-cellulose[1] and gel filtration (Seyer & Kang, 1978).

The peptide α1(III)-CB5 was further purified by rechromatography on CM-cellulose by the method previously described (Chung et al., 1974). Sixty-milligram samples were denatured at 45 °C for 15 min and applied to a 2.5 × 20 cm column of the resin previously equilibrated with 0.02 M sodium acetate, pH 4.8, at 43 °C. A concave gradient of NaCl from 0 to 0.12 M in the above acetate buffer with 830 mL of starting buffer and 500 mL of limiting buffer was used. The elution was obtained at 250 mL/h and the effluent was continuously monitored at 230 nm. Fractions containing the peptide α1(III)-CB5 were pooled, lyophilized, desalted on Sephadex G-25 in 0.1 M acetic acid, and relyophilized.

Amino Acid Analysis. Samples were hydrolyzed in constant boiling HCl at 105 °C for 24 h under an atmosphere of N_2. Analyses were performed on an automatic amino acid analyzer (Beckman Instruments, Palo Alto, CA) by using a single-column method previously described (Kang, 1972). No correction factors were used for losses of the labile amino acids or for the incomplete release of valine.

Enzymatic Hydrolysis. Digestions with trypsin (TPCK-treated, 3× crystallized, Worthington) were performed in 0.05 M NH_4HCO_3/0.005 $CaCl_2$, pH 8.0, at an enzyme/substrate ratio of 1:50 (Seyer & Kang, 1977). Thermolytic digestions were also performed in 0.05 M NH_4HCO_3, pH 8.3. Sixty milligrams of the peptide was dissolved in 12 mL of the buffer and allowed to react with 0.6 mg of thermolysin (Worthington) for 30 min at 37 °C. The reaction was terminated by lyophilization.

The rheumatoid synovial collagenase (Vater et al., 1978, 1979) was purified from the culture media of synovial explants by DEAE-cellulose chromatography, gel filtration on Ultrogel ACA-54 (Vater et al., 1979), and affinity chromatography on collagen Sepharose (Gellet et al., 1977) after activation with trypsin (TPCK-treated, Worthington). The mammalian collagenase was found to be protease-free as evidenced by its failure to cleave azocasein and by demonstration of the specific reaction products, α1(III)^A and α1(III)^B, at 34.5 °C, a temperature at which these peptides are denatured and rendered susceptible to the action of any contaminating, nonspecific proteases (Sakai & Gross, 1967). The reaction was monitored by measuring the specific viscosity of the collagen solution in a constant-temperature water bath in viscometers (Cannon Instruments). The reaction was stopped by the addition of EDTA to a final concentration of 0.005 M when the specific viscosity reached 40% of the initial specific viscosity. The presence of the specific reaction products was confirmed by $NaDodSO_4$ polyacrylamide gel electrophoresis in 7.5% acrylamide according to the method of Laemmli (1970). The collagen peptides were precipitated by the addition of solid ammonium sulfate to a final concentration of 25% (see below).

Hydroxylamine Cleavage. The cleavage of two Asn–Gly bonds of α1(III)-CB5 was achieved with hydroxylamine. The peptide α1(III)-CB5 (60 mg) was dissolved in 12 mL of H_2O and mixed with an equal volume of cold, freshly prepared 2 M NH_2OH in 1 M K_2CO_3 (pH 9.0) (Balian et al., 1971). The reaction was allowed to proceed for 30 min at 37 °C, terminated by adjusting the pH to 3.5 with HCl, and desalted by using a Sephadex G-25 column in 0.1 M acetic acid. The peptides were lyophilized and separated by Sephadex G-100 gel filtration (see below).

[1] Abbreviations used: CM-cellulose, carboxymethylcellulose; DEAE, diethylaminoethyl; EDTA, ethylenediaminetetraacetate; $NaDodSO_4$, sodium dodecyl sulfate; PhNCS, phenylthiohydantoin.

Edman Degradation. Automatic amino acid sequence analyses were performed by using a Beckman sequencer, Model 890C (Beckman Instruments, Palo Alto, CA), according to established principles (Edman & Begg, 1967). Both the Slow Protein–Quadrol (072172C) and the Slow Peptide–DMAA (071472) programs of Beckman Instruments were employed. Small peptides were treated with 2-amino-1,5-naphthalenedisulfonic acid in the presence of *N*-ethyl-*N'*-[3-(dimethylamino)propyl]carbodiimide to help retain the peptides in the reaction cup (Foster et al., 1973; Dixit et al., 1975). The phenylthiohydantoin amino acids were identified either by high-pressure liquid chromatography (Zimmerman et al., 1973) or after hydrolysis to their parent amino acids (Smithies et al., 1971). The COOH-terminal residues were, in most cases, not directly identified but inferred from the amino acid composition of the peptides and specificities of trypsin, hydroxylamine, or CNBr. Subsequent overlapping sequence analyses confirmed their identify.

Column Chromatography. Peptides obtained after trypsin, thermolysin, or hydroxylamine cleavage of α1(III)-CB5 were separated by a combination of gel filtration, ion-exchange chromatography, and reverse-phase chromatography. The tryptic peptides were initially fractionated by Sephadex G-50S gel filtration (4.0 × 110 cm column) in 0.1 M acetic acid. Smaller peptides (3–6 residues) were further separated by automated peptide analysis on a 0.9 × 25 cm column of PA-35 resin (Beckman Instruments, Palo Alto, CA) at 60 °C (Kang & Gross, 1970). An automatic analyzer (Technicon Instruments, Inc., Ardsley, NY) equipped with a stream-splitting device allowed continuous monitoring of a portion of the effluent for ninhydrin reactivity. The remaining 90% of the column effluent was collected in a fraction collector. Separation was achieved by utilizing a nine-chamber gradient starting with 0.02 M sodium citrate buffer, pH 3.8 (Technicon peptide methodology). Fractions were collected, their pH was adjusted to 2.0 with 1 N HCl, and they were desalted by using 1 × 2 cm columns of Aminex 50-X8 (200–400 mesh) (Bio-Rad Laboratories, Richmond, CA) (Hirs, 1967). Intermediate-size tryptic peptides (7–15 residues) were separated by phosphocellulose chromatography (P-11 Whatman). The 1 × 6 columns of phosphocellulose were equilibrated with 0.001 M sodium acetate, pH 3.8, at 43 °C. Samples in 5 mL of buffer were applied and elution was obtained by using a linear gradient of NaCl from 0 to 0.1 M NaCl over a total volume of 600 mL (flow rate 60 mL/h). Larger peptides (17–24 residues) were separated by high-pressure liquid chromatography (Waters Associates, Milford, MA) (Rivier, 1978) by using an ODS-2 Magnum 9 reverse-phase column (Whatman, Inc., Bidwell, NJ). The initial solvent (A) consisted of 5% acetonitrile (Burdick and Jackson) and 5% 0.01 M K_2HPO_4, pH 7.0, and the limiting solvent contained 50% acetonitrile and 50% 0.01 M K_2HPO_4, pH 7.0. A flow rate of 2.5 mL/min was maintained with a linear gradient (program 6) of the solvent programmer (Waters Associates, Milford, MA) going from 6% B to 85% B in 45 min. Peptides were continuously monitored by using a variable wavelength spectrophotometer. The two largest peptides (27 and 33 residues) were isolated from the first peak of Sephadex G-50S by ion-exchange chromatography on phosphocellulose as above except with 0.01 M sodium acetate, pH 4.8, as the intial buffer and a NaCl gradient of 0.0–0.3 M (total volume 800 mL).

The thermolytic reaction products were fractionated initially by CM-cellulose chromatography. The digestion mixture was dissolved in 50 mL of 0.01 M sodium acetate, pH 4.8, and applied to a 2.5 × 20 cm column of CM-cellulose (43 °C)

Table I: Amino Acid Composition of α1(III)-CB5 and Its Tryptic Peptides from Human Cirrhotic Liver[a]

amino acid	α1(III)-CB5	T1	T2	T3	T4	T5[b]	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	total
Hyp[c]	30.1	1.1		0.9	1.1		0.9	3.1	0.8	0.9	1.8	2.6	3.5	3.7	2.6	2.4	3.2	
	(30)	(1)		(1)	(1)		(1)	(3)	(1)	(1)	(2)	(3)	(4)	(4)	(3)	(3)	(4)	32
Asp		0.2	0.8				0.2	0.2	1.2	0.1	2.3	2.3	1.2	2.3	0.2	0.2	1.4	
	(10)		(1)						(1)		(2)	(2)	(1)	(2)			(1)	10
Thr	3.8							0.1	0.2	0.2	0.2	0.8	0.9	0.2	0.9		0.9	
	(4)											(1)	(1)		(1)		(1)	4
Ser				1.1				1.0	2.1	0.8		2.2	0.1	1.7	0.7	0.8	0.2	
	(10)			(1)				(1)	(2)	(1)		(2)		(2)	(1)	(1)		11
Glu		1.3		0.9	0.8	1.8	0.1	0.2		0.2	1.0	0.2	3.2	0.3	1.2	2.1	2.4	
	(14)	(1)		(1)	(1)	(2)					(1)		(3)		(1)	(2)	(2)	14
Pro[c]	27.8	0.2	0.7				1.2	0.4	0.2	1.1	1.4	3.4	2.4	4.4	2.3	5.8	5.7	25
	(28)		(1)				(1)			(1)	(1)	(3)	(2)	(4)	(2)	(5)	(5)	
Gly		2.3	1.1	2.1	1.2	2.3	3.3	6.3	3.3	3.9	7.2	7.1	9.2	8.4	8.2	10.2	12.3	
	(85)	(2)	(1)	(2)	(1)	(2)	(3)	(6)	(3)	(4)	(7)	(7)	(9)	(8)	(8)	(10)	(12)	85
Ala		1.0					0.9	3.1	1.1		1.2	1.9	2.0	0.9	4.2	2.3	3.0	
	(21)	(1)					(1)	(3)	(1)		(1)	(2)	(2)	(1)	(4)	(2)	(3)	21
Val	3.6										0.8	0.6			0.2	1.7		
	(4)										(1)	(1)				(2)		
Ile	4.2												0.2	0.7	1.8	0.1	1.8	
	(4)													(1)	(2)		(2)	4
Leu	4.9						0.8	0.1		1.0				1.1	0.9	0.3	1.1	
	(5)						(1)			(1)				(1)	(1)		(1)	5
Phe	1.8							0.9					0.9					
	(1)							(1)					(1)					2
Hyl[c]	1.1										0.2	0.3	0.1			0.7	0.2	
	(1)															(1)		1
Lys[c]	9.2	1.1						0.1	1.1	0.8	2.1	0.8	0.9	1.1	0.1	0.4	1.1	
	(9)	(1)							(1)	(1)	(2)	(1)	(1)	(1)			(1)	9
Arg	8.2		1.1	0.9	1.0	2.0		1.1							1.2		1.0	
	(8)		(1)	(1)	(1)	(2)		(1)							(1)		(1)	8
Hse[d]	1.1						1.1											
	(1)						(1)											
total	237	6	4	6	4	6	8	15	9	9	17	22	24	23	24	27	33	237

[a] Values expressed as residues per peptide. No entry indicates the level was less than 0.1 residue per peptide. Integral values were given for residues present in amounts in greater than 10 residues. Numbers in parentheses indicate assumed integral values. [b] The peptide Gly-Glu-Arg occurs twice and hence molar ratios of this peptide were twofold over the other peptides isolated by identical procedures. [c] Partial proline and lysine hydroxylation was noted by sequence analysis. This accounts for the frequent nonintegral values obtained by amino acid analysis and the discrepancy between amino acid content of α1(III)-CB5 when analyzed by amino acid analysis and by sequence studies. [d] Includes homoserine lactone.

equilibrated with the same buffer. Separation was achieved by utilizing the above 0.01 M sodium acetate buffer with a linear 0.0–0.06 M NaCl gradient over a total volume of 2 L. Each peptide fraction was further purified by Sephadex G-50S gel filtration in 0.1 M acetic acid. In certain cases, isolated thermolytic peptides were cleaved with trypsin. The tryptic cleavage products of the thermolytic peptides were separated by Sephadex G-50S gel filtration.

The hydroxylamine cleavage products were separated initially by Sephadex G-100 gel filtration in 0.05 M Tris/1 M $CaCl_2$, pH 7.4, by using a 3.0 × 110 cm column. Fifteen-milliliter fractions were collected and the major UV-absorbing peaks were further subjected to CM-cellulose chromatography. A 2.5 × 20 cm column equilibrated with 0.02 M sodium acetate, pH 4.8, at 43 °C with a linear gradient of 0.0–0.12 M NaCl (total volume 1600 mL) was used to separate the fragments. Each peptide was further purified by Sephadex G-50S gel filtration in 0.1 M acetic acid for amino acid analysis and sequential degradation.

Collagenase cleavage products of native type III collagen were separated from the enzyme by 25% $(NH_4)_2SO_4$ precipitation and three subsequent washings. The insoluble residue was dialyzed against 0.1 M acetic acid at 4 °C and lyophilized. The dried residue was reduced with 0.1 M dithiothreitol in 5 M guanidine hydrochloride at pH 7.4 (40 °C) for 4 h under an atmosphere of N_2. Separation of $\alpha1(III)^A$ and $\alpha1(III)^B$ was accomplished by gel filtration of the reduced product on a 2.0 × 110 column of agarose A1.5M equilibrated with 2.5 M guanidine hydrochloride/0.05 M Tris, pH 7.4, containing 0.01 M dithiothreitol. The effluent was monitored at 232 nm and the 24000 mol wt fragment $\alpha1(III)^B$ was collected, desalted on Sephadex G-25 in 0.1 M acetic acid, and lyophilized.

Results

Isolation of Tryptic Peptides. The α1(III)-CB5 peptide was purified from CNBr digests of type III collagen of human cirrhotic liver as previously described by using CM-cellulose chromatography. Tryptic digestion yielded a mixture of peptides which could be fractionated into four apparent molecular weight sizes by Sephadex G-50S gel filtration.[2] Five smaller tryptic peptides, T1–T5,[3] containing 3–6 residues were isolated by PA-35 ion-exchange chromatography.[2] T-5 consisted of 2 molar equiv of Gly-Glu-Arg and was later located in two positions of α1(III)-CB5 (Table I). The intermediate-size peptides (fraction 3 of Sephadex G-50S) contained tryptic peptides with 6–15 residues. These were separated by phosphocellulose chromatography into four peptides,[2] T6–T9 (Table I), and a small amount of the hexapeptide, T1, isolated above. T6 was an octapeptide with a residue of homoserine and, therefore, must be the COOH-terminal tryptic peptide.

[2] See paragraph at end of paper regarding supplementary material.

[3] The peptides were designated numerically on the basis of their elution position from the specific chromatographic separation. T stands for tryptic peptides, Th for thermolytic peptides, and HA for hydroxylamine-derived peptides.

The double peak seen for T6 on phosphocellulose chromatography was consistent with the presence of homoserine and homoserine lactone, as has been observed previously (Seyer & Kang, 1978). The larger tryptic peptides from Sephadex G-50S gel filtration (fraction 2) (15–24 residues) were somewhat more difficult to separate by ion-exchange chromatography but were subsequently separated by reverse-phase high-pressure liquid chromatography[2] into five peptides (T10–14) (Table I). Finally, the two largest tryptic peptides, T15 and T16, were isolated from fraction 1 of Sephadex G-50S by phosphocellulose chromatography at pH 4.8.[2] The amino acid composition of these tryptic peptides is presented in Table I. The sum of their amino acid contents accounts for the amino acid content of the intact α1(III)-CB5 within experimental error.

Since there is a total of 18 basic residues (8 Arg, 9 Lys, and 1 Hyl), one would have expected 19 tryptic peptides. However, T10 contained two residues of Lys, and T16 contained a residue each of Lys and Arg, the Arg residue being the COOH terminus. In both peptides, one of the Lys residues occurred in the sequence of Gly-Asp-Lys-Gly and was not cleaved by trypsin. The reason for the resistance of the lysyl residues in these two instances is not clear, but similar resistance was previously encountered in other collagenous peptides when Asp preceded Lys in sequence (Seyer & Kang, 1978). As stated previously, T5 contained 2 equiv of Gly-Glu-Arg, suggesting a repeat of this sequence. Thus in summary, the 16 peptides presented in Table I account for the expected tryptic peptides of α1(III)-CB5. Amino acid sequence analysis of each tryptic peptide confirmed its composition.

Alignment of Tryptic Peptides. Alignment of the tryptic peptides in α1(III)-CB5 was deduced from automated Edman degradation of intact α1(III)-CB5, thermolytic peptides, hydroxylamine peptides, the mammalian collagenase cleavage product, α1(III)^B, and the sequence of the individual tryptic peptides. First, sequential degradation of intact α1(III)-CB5 through the first 54 residues from the NH_2 terminus provided the necessary information to align the tryptic peptides T4-T9-T10-T2-T16 (Figure 1).

Next, additional information was obtained from characterization of thermolytic peptides of α1(III)-CB5. A thermolytic digest of α1(III)-CB5 was separated by CM-cellulose chromatography which yielded nine peptides.[2] Of these, only the peptides Th3, -7, -8, and -9 were needed to help establish the alignment. Th7, by amino acid analysis (Table II) and sequential degradation (Figure 1), represented the NH_2-terminal peptide containing residues 1–62. After tryptic hydrolysis of Th7, a peptide of 28 residues (Th7-1) was isolated by Sephadex G-50S chromatography. A complete sequence analysis of Th7-1 provided confirmation of T16 and its location as the COOH terminus of Th7. Another thermolytic peptide, Th8, was purified by Sephadex G-50S chromatography. Its amino acid composition (Table II) and sequence analysis (Figure 1) indicated it to be residues 63–113. Its NH_2-terminal sequence was Ile-Ala-Gly-Pro-Arg (residues 63–67), consistent with the COOH-terminal portion of T16 (Figure 1). Edman degradation of Th8 through 27 cycles allowed the alignment of T16 (partial)-T3-T12. Isolation of T12 from the tryptic digestion of Th8 by Sephadex G-50S confirmed its location in Th8.

Further information was obtained from characterization of hydroxylamine fragments of α1(III)-CB5. Hydroxylamine cleavage products were initially fractionated on Sephadex G-100.[2] A major portion of the material eluted in the region of 7000–8000 mol wt, and, from this, three peptides were

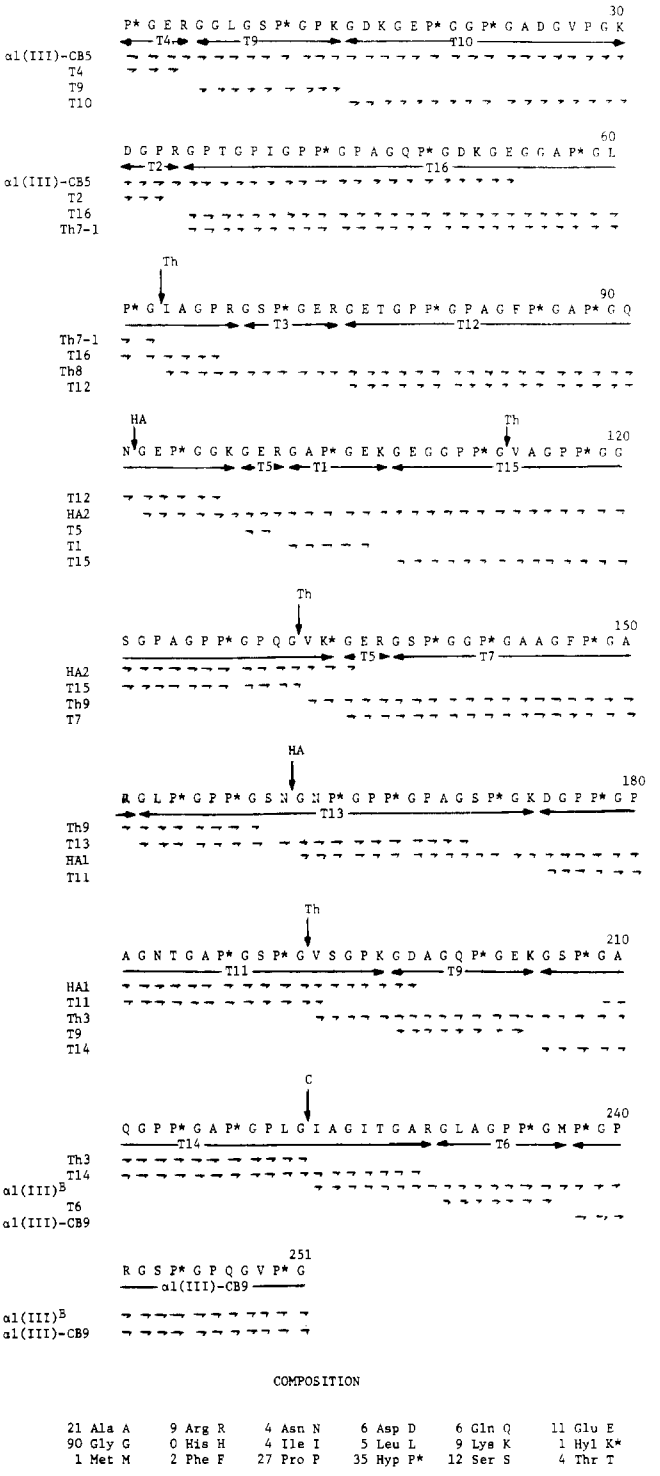

FIGURE 1: The complete amino acid sequence of α1(III)-CB5 of human liver type III collagen. The tryptic peptides isolated are indicated by long arrows (↔). Short arrows (—) indicate the extent of Edman degradation of each peptide.

purified.[2] However, information derived from investigation of two of them, HA1 and HA2, was sufficient to align the tryptic peptides. Amino acid composition (Table II) and automated Edman degradation (Figure 1) of HA2 through 43 cycles indicated it to consist of residues 92–160. Its NH_2-terminal sequence contained the COOH-terminal portion of T12, which is followed by T5, T1, and T15, in that order.

That T5, T7, and T13, in that order, follow T15 in alignment was deduced from characterization of another thermolytic peptide, T9. From its amino acid composition (Table II)

Table II: Amino Acid Composition of Thermolytic and Hydroxylamine-Derived Peptides of α1(III)-CB5 of Human Cirrhotic Liver[a]

amino acid	position[b] 192–221 Th3	1–62 Th7	63–113 Th8	132–191 Th9	162–237 HA1	92–161 HA2
4-Hyp[c]	3.5 (4)	8.1 (8)	7.1 (7)	11.4 (11)	10.9 (11)	10.1 (10)
Asp	9.9 (1)	4.4 (4)	1.3 (1)	2.9 (3)	3.4 (3)	0.8 (1)
Thr	0.2	1.3 (1)	1.0 (1)	0.9 (1)	2.4 (2)	-
Ser	1.6 (2)	0.7 (1)	0.9 (1)	3.6 (4)	4.1 (4)	3.2 (3)
Glu	3.3 (3)	4.1 (4)	7.3 (7)	0.9 (1)	2.7 (3)	5.8 (6)
Pro	3.0 (3)	7.4 (7)	4.4 (4)	4.9 (5)	8.2 (8)	5.0 (5)
Gly	10 (10)	24 (24)	19 (19)	20 (20)	26 (26)	27 (27)
Ala	2.7 (3)	3.3 (3)	4.0 (4)	5.8 (6)	9.0 (9)	6.3 (6)
Val	0.6 (1)	0.5 (1)	-	0.7 (1)	0.8 (1)	1.5 (2)
Ile	-	0.7 (1)	0.9 (1)	-	2.3 (2)	-
Leu	0.9 (1)	2.3 (2)	0.2	0.1 (1)	2.1 (2)	0.7 (1)
Phe	-	-	0.8 (1)	1.1 (1)	0.2	0.8 (1)
Hyl	0.3	0.2	0.3	0.9 (1)	0.5	1.2 (1)
Lys	1.9 (2)	4.3 (4)	1.8 (2)	1.0 (1)	2.8 (3)	1.7 (2)
Arg	-	1.8 (2)	3.1 (3)	2.0 (2)	0.9 (1)	3.3 (3)
Hse[d]			0.2		0.8 (1)	0.3
total	30	62	51	58	76	68

[a] Values expressed as residues per peptide. A dash (-) represents less than 0.1 residue per peptide. Integral values were given for residues present in greater than 10 residues. Numbers in parentheses indicate assumed integral values. [b] Location of the peptide in α1(III)-CB5. [c] Partial proline and lysine hydroxylation was noted by sequence analysis. Actual values were therefore given for these amino acids. [d] Includes homoserine lactone.

and sequence analysis (Figure I) through 27 cycles, it was determined that it is a peptide of 60 residues being at residue 132. The NH_2-terminal sequence of Th9, Val-Hyl-Gly, overlapped a portion of HA2 and the COOH terminus of T15. Since the sequence of Val-Hyl occurred only once in the entire α1(III)-CB5, the order of T15-T5-T7-T13 could be deduced without ambiguity. Characterization of HA1 (Table II) (residues 161–237) and its sequence determination through the first 38 residues from the NH_2 terminus allowed establishment of the alignment of T11 to follow T13. Finally, isolation and sequence determination of another thermolytic peptide, Th3 (residues 192–221) (Table II), established the position of T9 and T14. T6 must be the COOH terminus since it contains a residue of homoserine. Further evidence substantiating this alignment was obtained by the use of mammalian collagenase. Native type III collagen was digested with purified rheumatoid synovial collagenase, and the resulting products, $\alpha1(III)^A$ and $\alpha1(III)^B$, were separated by gel filtration on agarose A5M.[2] The sequence of the COOH fragment, $\alpha1(III)^B$, was determined by beginning with the cleavage site and was found to be identical with the COOH-terminal portion of T14 (Figure 1), providing overlapping sequences of T14-T6. Furthermore, sequence analysis of $\alpha1(III)^B$ through the methionyl residue (residue 237) into the subsequent CNBr peptide, CB9, provided confirmation of the order of the CNBr peptides, α1(III)-CB5-9.

Thus, in summary, the alignment of tryptic peptides in α1(III)-CB5 is T4-9-10-2-16-3-12-5-1-15-5-7-13-11-9-14-6.

Internal Sequence of Various Peptides. All peptides were modified at the COOH groups by treatment with 2-amino-1,5-naphthalenedisulfonic acid in the presence of *N*-ethyl-*N'*-[3-(dimethylamino)propyl]carbodiimide. The complete amino acid sequence of α1(III)-CB5 obtained in the present study is summarized in Figure 1. Also indicated in the figure are the various peptides used to arrive at the sequence and the specific cycles of residues degraded (short horizontal half-arrows). Only the minimal data essential for the deduction of the sequence are actually shown in the figure. Additional data including yields of the automated Edman degradation of peptides are summarized in the supplementary material.[2]

Discussion

Type III collagen was obtained by limited pepsin digestion of human cirrhotic liver. After selective salt precipitation and purification, the collagen was cleaved with CNBr. Ten distinct peptides were obtained, ranging from 3 to 246 residues. The primary structure of eight CNBr peptides, representing the NH_2-terminal half of the α1(III) chain, have been determined (Seyer & Kang, 1977, 1978). In the present report, the primary structure of another peptide, α1(III)-CB5, representing 237 amino acids, has been determined, thus establishing the amino acid sequence of 795 amino acid residues of human cirrhotic liver type III collagen. The use of collagenase digestion of native type III collagen gave a sequence overlap with α1(III)-CB9, the COOH-terminal peptide of the α1(III) chain.

The covalent-structure study was first initiated through isolation of the 16 tryptic peptides. Two peptides contained Gly-Asp-Lys-Gly sequences (T10, T16) which were isolated as uncleaved tryptic peptides. Both phenylalanine residues were followed by Hyp, thus restricting the use of chymotrypsin since previous studies have shown Phe–Hyp bonds to be resistant to enzymatic cleavage. The two Asn–Gly bonds were cleaved with hydroxylamine. Since only a partial cleavage of these bonds is generally obtained (Balian et al., 1971), the theoretical number of possible peptides would be large. No attempts were made to quantitate the amount of cleavage or to isolate all peptides aside from HA1, HA2, and HA3. HA1 and HA2 contained NH_2-terminal sequences consistent with hydroxylamine cleavage at positions 91–92 and 160–161. HA3 was the NH_2 terminus of α1(III)-CB5.

Thermolytic digestion also produced a large number of peptides. The major cleavage sites found were Gly-Ile (62–63), Gly-Val (113–114), Gly-Val (131–132), and Gly-Val (191–192). The sequence Gly-Ala (144–145) was also a minor cleavage site (not used here). Other theoretical cleavage points, Gly-Ile (221–222), Gly-Ile (224–225), and Gly-Leu (230–231), could not be identified.

The use of purified rheumatoid synovial collagenase was useful in obtaining confirmational evidence of the alignment and structure of two tryptic peptides, T14 and T6, at the COOH terminus of α1(III)-CB5. Native type III collagen was digested with purified synovial collagenase, and the reduced cleavage products, $\alpha1(III)^A$ and $\alpha1(III)^B$, were isolated by gel filtration. Automated amino acid sequence analysis of $\alpha1(III)^B$ enabled identification of the sequence of Ile-Ala-Gly-Ile-Thr-Gly-Ala-Arg-Gly-Leu-Ala-Gly-Pro-Hyp-Gly-Met-Hyp-Gly-Pro-Arg-Gly. . ., confirming the COOH-terminal sequence of T14, the overlapping sequence with T6, and the overlapping sequence with the next CNBr peptide of type

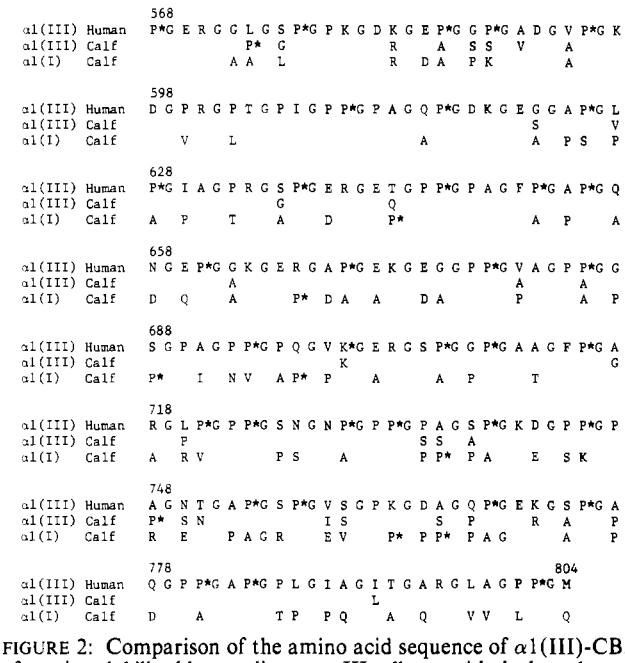

FIGURE 2: Comparison of the amino acid sequence of α1(III)-CB5 of pepsin-solubilized human liver type III collagen with the homologous regions from calf α1(III)-CB5 and calf α1(I)-CB7 from their respective collagen chains. A methionine at residue 567 of the calf α1(I) collagen chain and the corresponding positions of both calf and human liver α1(III) collagen chains enabled isolation of CNBr peptides initiating with position 568 and continuing through residue 804 with α1(III) collagen chains. No methionine is present at position 804 of the calf α1(I) collagen chain. Identical residues in the same position of each chain are indicated by blank spaces. The position numbers are based on calf α1(I) collagen-chain residue numbers.

III collagen, α1(III)-CB9. This is consistent with the report (Chung et al., 1974) that α1(III)-CB9 was the COOH-terminal CNBr peptide of human α1(III).

It has been reported that the Arg residue (229) represents the single trypsin-sensitive bond in the native type III collagen molecule (Miller et al., 1976a). The collagenase-specific cleavage site found here was the Gly–Ile bond (221–222). This is at variance with a previous report (Miller et al., 1976b) that the collagenase cleavage site was a Gly–Leu bond. Since our report utilized human cirrhotic liver type III collagen as compared to human skin type III collagen previously reported, the possibility that tissue heterogeneity and/or microheterogeneity of type III collagen similar to that found with type II cartilage collagen (Butler et al., 1977a,b) might be responsible for the discrepancy cannot be excluded. The comparison of the collagenase-specific cleavage site of type I collagen (Gross et al., 1974) and human liver type III collagen presented here revealed that type III collagen was cleaved one triplet proximal to the cleavage found in type I collagen. This agrees with a previous study (Miller et al., 1976b) that the cleavage site is 16 residues from the COOH terminus of α1-(III)-CB5, but a discrepancy was found in the Met between α1(III)-CB5 and α1(III)-CB9. We found this to be one triplet proximal to that which was previously suggested. A second Gly-Ile-Ala sequence also occurs in α1(III)-CB5 (residue 62–64), but native type III collagen apparently is not cleaved at this position by mammalian collagenase, confirming the high specificity of the enzyme (Gross et al., 1974; Highberger et al., 1979).

The remaining sequence of α1(III)-CB5 is similar in many respects to other collagen primary structures thus far identified (Hulmes et al., 1973; Gallop & Pas, 1975; Fietzek & Kühn, 1976; Piez, 1976). Gly occurs at every third residue, and Hyp and Hyl are restricted to the Y position of the Gly-X-Y triplet. In agreement with the excess Gly content of type III collagen, Gly occurs in both X and Y positions. Gly-Gly-Hyp appeared most frequent (three times). Other sequences include Gly-Gly-Leu, Gly-Gly-Lys, and Gly-Glu-Gly (twice each). Thus far with human type III collagen, Gly has been found in the X position 18 times, 6 of these being Gly-Gly-Hyp, indicating a preference for this triplet. Four of the remaining were the Gly-Gly-Lys triplet, and four Gly were in the Y position. No such preferences were noted for Gly in the Y position. Proline in the Y position was generally only partially hydroxylated. Residues 28 and 163 were estimated to be less than 20% hydroxylated for unknown reasons. Lysine at 133 was fully hydroxylated. Other lysine residues were less than 20% hydroxylated.

The present data also allow a comparison of the homologous sequence from α1(I)-CB7 and α1(III)-CB5 of calf skin (Figure 2). Only 29 substitutions, all conservative, were found between α1(III)-CB5 of human and α1(III)-CB5 of calf. This is similar to the level of identity (89%) found between chick and calf α1(I) peptides (Dixit et al., 1977). Interestingly, 13 of these substitutions involve threonine or serine substitutions to another amino acid. The reason for this instability at the genetic level is obscure. The homologous region of α1(I)-CB7 of calf skin (Fietzek et al., 1979) contains 86 substitutions (63% identity). This is similar to the degree of variation between other portions of the α1(III) and α1(I) chains thus far examined and between the α1(I) chain and α2 chains (Seyer & Kang, 1977, 1978). Substitution between α1-(III)-CB5 of human and α1(I)-CB7 involved numerous minor substitutions (no alteration is net charge) and also changes of basic residues to neutral amino acids (six cases) and acidic residues to neutral amino acids (four cases). This may account for alterations in the banding pattern found between SLS aggregates of type I and III collagen (Trelstad et al., 1974) when examined by electron microscopy.

In summary, the amino acid sequence of 795 residues of human type III collagen has been determined thus far. The conservative interspecies substitutions between the human and calf type III collagen are apparent, but greater variances are found when comparing α1(III) and α1(I) chains.

Acknowledgments

The authors are grateful to Dr. J. M. Young of this Veterans Administration Medical Center for pathological examination of tissue samples and for the excellent technical assistance of E. Rowland, E. Ryan, D. Blanton, T. Todd, and A. Crosby.

Supplementary Material Available

Nine figures (isolation and characterization of peptides by various chromatographic methods) and one table (quantitation of phenylthiohydantoin derivatives after automatic Edman degradation of peptides from human liver type III collagen) (10 pages). Ordering information is given on any current masthead page.

References

Allmann, H., Fietzek, P. P., Glanville, R. W., & Kühn, K. (1979) *Hoppe-Seyler's Z. Physiol. Chem. 360*, 861.

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

Bentz, H., Fietzek, P. P., & Kühn, K. (1979) *Hoppe-Seyler's Z. Physiol. Chem. 360*, 833.

Butler, W. T., Finch, J. E., Jr., & Miller, E. J. (1977a) *J. Biol. Chem. 252*, 639.

Butler, W. T., Miller, E. J., & Finch, J. E., Jr. (1977b) *Biochemistry 16*, 4981.

Chung, E., & Miller, E. J. (1974) *Science 183*, 1200.
Chung, E., Keele, E. M., & Miller, E. J. (1974) *Biochemistry 13*, 3459.
Dewes, H., Fietzek, P. P., & Kühn, K. (1979a) *Hoppe-Seyler's Z. Physiol. Chem. 360*, 821.
Dewes, H., Fietzek, P. P., & Kühn, K. (1979b) *Hoppe-Seyler's Z. Physiol. Chem. 360*, 851.
Dixit, S. N., Seyer, J. M., Oronskey, A. P., Corbett, C., Kang, A. H., & Gross, J. (1975) *Biochemistry 14*, 1933.
Dixit, S. N., Seyer, J. M., & Kang, A. H. (1977) *Eur. J. Biochem. 81*, 599.
Edman, P., & Begg, G. (1967) *Eur. J. Biochem. 1*, 80.
Epstein, E. H., Jr. (1974) *J. Biol. Chem. 249*, 3225.
Fietzek, P. P., & Kühn, K. (1976) *Int. Rev. Connect. Tissue Res. 7*, 1.
Fietzek, P. P., Allmann, H., Rauterberg, J., Henkel, W., Wachter, E., & Kühn, K. (1979) *Hoppe-Seyler's Z. Physiol. Chem. 360*, 890.
Foster, J. A., Bruenger, C. L., Hu, C. L., Albertson, K., & Franzblau, C. (1973) *Biochem. Biophys. Res. Commun. 53*, 70.
Gallop, P. M., & Paz, M. A. (1975) *Physiol. Rev. 55*, 418.
Gallop, P. M., Blumenfield, O. O., & Seifter, S. (1972) *Annu. Rev. Biochem. 41*, 617.
Gellet, C., Eeckhout, Y., & Vaes, G. (1977) *FEBS Lett. 74*, 126.
Gross, J., Harper, E., Harris, E. D., Jr., McCroskery, P. A., Highberger, J. H., Corbett, C., & Kang, A. H. (1974) *Biochem. Biophys. Res. Commun. 61*, 605.
Highberger, J. H., Corbett, C., & Gross, J. (1979) *Biochem. Biophys. Res. Commun. 89*, 202.
Hirs, C. H. W. (1967) *Methods Enzymol. 11*, 386.
Hulmes, D. J. S., Miller, A., Parry, D. A., Piez, K. A., & Woodhead-Galloway, J. (1973) *J. Mol. Biol. 70*, 137.
Kang, A. H. (1972) *Biochemistry 11*, 1828.
Kang, A. H., & Gross, J. (1970) *Biochemistry 9*, 796.
Laemmli, V. (1970) *Nature (London) 227*, 680.
Lang, H., Glanville, R. W., Fietzek, P. P., & Kühn, K. (1979) *Hoppe-Seyler's Z. Physiol. Chem. 360*, 841.
Miller, E. J., & Lunde, L. G. (1973) *Biochemistry 12*, 3153.
Miller, E. J., Finch, J. E., Chung, E., & Butler, W. T. (1976a) *Arch. Biochem. Biophys. 173*, 631.
Miller, E. J., Harris, E. D., Jr., Chung, E., Finch, J. E., Jr., McCroskery, P. A., & Butler, W. T. (1976b) *Biochemistry 15*, 787.
Piez, K. A. (1976) in *Biochemistry of Collagen* (Ramachandran, G. N., & Reddi, A. H., Eds.) Plenum, New York.
Rivier, J. E. (1978) *J. Liquid Chromatogr. 1*, 343.
Sakai, T., & Gross, J. (1967) *Biochemistry 6*, 518.
Seyer, J. M., & Kang, A. H. (1977) *Biochemistry 16*, 1158.
Seyer, J. M., & Kang, A. H. (1978) *Biochemistry 17*, 3404.
Seyer, J. M., Hutcheson, E. T., & Kang, A. H. (1976) *J. Clin. Invest. 57*, 1498.
Seyer, J. M., Hutcheson, E. T., & Kang, A. H. (1977) *J. Clin. Invest. 59*, 241.
Smithies, O., Gibson, D., Fanning, E. M., Goodfliesh, R. M., Gilman, J. G., & Ballanthyne, D. L. (1971) *Biochemistry 10*, 4912.
Stuart, J. M., Cremer, M. A., Dixit, S. N., Kang, A. H., & Townes, A. S. (1979) *Arthritis Rheum. 22*, 347.
Swann, D. A., Constable, I. J., & Harper, E. (1972) *Invest. Ophthalmol. 11*, 735.
Traub, W., & Piez, K. A. (1971) *Adv. Protein Chem. 25*, 243.
Treslstad, R. L. (1974) *Biochem. Biophys. Res. Commun. 57*, 717.
Vater, C. A., Mainardi, C. L., & Harris, E. D., Jr. (1978) *Biochim. Biophys. Acta 539*, 238.
Vater, C. A., Mainardi, C. L., & Harris, E. D., Jr. (1979) *J. Biol. Chem. 254*, 3054.
Zimmerman, C. L., Pisano, J. J., & Appella, E. (1973) *Biochem. Biophys. Res. Commun. 55*, 1220.