Covalent Structure of Collagen: Amino Acid Sequence of α1(III)-CB9 from Type III Collagen of Human Liver†

Jerome M. Seyer* and Andrew H. Kang

ABSTRACT: The peptide α1(III)-CB9 was prepared and purified from human liver, and its amino acid sequence was determined. Automated Edman degradation of the intact peptide and peptides derived from selective cleavage with hydroxylamine and digestions with trypsin, thermolysin, and Staph V8 protease enabled determination of the complete amino acid sequence. The peptide α1(III)-CB9 represents the COOH terminus of the helical (pepsin-resistant) portion of type III collagen and terminates in a Cys-Cys sequence responsible for the intramolecular disulfide cross-linkages with other chains. The present work completes the entire amino acid sequence of the helical (pepsin-resistant) portion of human cirrhotic liver type III collagen consisting of peptides α1-(III)-CB3-7-6-1-8-10-2-4-5-9. The COOH terminus of human liver α1(III) contained two additional triplets which, together with the extra triplet at the NH_2 terminus in α1(III)-CB3, make the helical portion of type III collagen longer than α1(I) by nine residues (three Gly-X-Y triplets). The helical region of human liver type III collagen, therefore, consists of 1023 amino acids or 341 triplets.

Ten CNBr peptides of type III collagen from human liver have been isolated and the amino acid sequence of nine of the peptides, α1(III)-CB3-7-6-1-8-10-2-5 representing the initial 795 residues of the NH_2 terminus has been reported (Seyer & Kang, 1977, 1978; Seyer et al., 1980). The present report describes the amino acid sequence of α1(III)-CB9, a 236-residue peptide located at the COOH terminus of the helical portion of type III collagen which contains the disulfide intramolecular cross-link (Chung & Miller, 1974). This, therefore, concludes the amino acid sequence analysis of the helical part of human liver type III collagen. While this work was being completed, the entire sequence of the helical portion of type III collagen of bovine skin was reported (Fietzek et al., 1979; Dewes et al., 1979a,b; Bentz et al., 1979; Lang et al., 1979; Allman et al., 1979), allowing a comparison of the complete helical structure of type III collagen from the two species.

Materials and Methods

Preparation of α1(III)-CB9. Human cirrhotic livers were obtained after autopsy, and type III collagen was prepared as previously described (Seyer et al., 1977). The CNBr peptides were obtained by digestion of purified type III collagen in 70% formic acid at 40 °C and separated by ion-exchange chromatography on CM-cellulose[1] and by gel filtration (Seyer & Kang, 1977).

The peptide, α1(III)-CB9, was eluted from the CM-cellulose column by using a final column wash with 0.5 M NaCl in 0.02 M sodium citrate, pH 3.83, at 43 °C. The peptide α1-(III)-CB9 was further purified by isolation of the 66 000 mol wt trimer by gel filtration on agarose A1.5m with 0.05 M Tris/1 M $CaCl_2$, pH 7.4, as the eluting solvent. The peptide (80 mg) was dissolved in 10 mL of eluting solvent, denatured at 45 °C for 15 min, and applied to a 4.0 × 110 cm column at 22 °C; the effluent was continuously monitored at 230 nm with a Gilford Model 2400S recording spectrophotometer. Fractions corresponding to 66 000 mol wt were collected, desalted on a Sephadex G-25 column (4.0 × 60 cm) in 0.1 M acetic acid, and lyophilized. In some instances, α1(III)-CB9 was prepared by agarose A1.5m chromatography of the entire CNBr digest, since no other peptide of this size (66 000) occurs in type III collagen CNBr digests.

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) using a single-column method previously described (Kang, 1972). Cysteine was analyzed as the *S*-carboxymethyl derivative. No correction factors were used for losses of labile amino acids or incomplete release of valine or isoleucine.

Hydroxylamine (HA) Cleavage. The cleavage of sensitive Asn–Gly bonds was achieved with hydroxylamine. The peptide, α1(III)-CB9 (120 mg), was dissolved in 25 mL of freshly prepared 1 M NH_2OH in 0.5 M K_2CO_3 (pH 9.0) (Balian et al., 1971). The reaction was allowed to proceed for 30 min at 45 °C, terminated by adjusting the pH to 3.5 with HCl, and desalting by using a Sephadex G-25 column in 0.1 M acetic acid. The peptides were lyophilized and separated on Sephadex G-75S in 0.04 M sodium acetate, pH 4.8 (see below).

Enzymatic Hydrolysis. Digestions with trypsin (TPCK treated, 3X crystallized, Worthington Biochemical Corp.) were performed in 0.2 M NH_4HCO_3/0.001 M $CaCl_2$, pH 8.0, at an enzyme (E)/substrate (S) (w/w) ratio of 1:50 (Seyer & Kang, 1977). Thermolytic (Worthington) digestions (E/S = 1:200) and Staph V8 protease (Miles Laboratories) digestions (E/S = 1:20) were performed in 0.02 M NH_4HCO_3, pH 8.0, at 37 °C, by using 45-min incubations for the former and 16 h for the latter enzyme. In all cases, the reaction was terminated by lyophilization.

Edman Degradation. Automatic amino acid sequence analyses were performed by using a Beckman Sequencer, Model 890C, (Beckman Instruments, Palo Alto, CA), ac-

† From the Veterans Administration Medical Center (J.M.S.) and the Departments of Biochemistry and Medicine (A.H.K.), University of Tennessee Center for the Health Sciences, Memphis, Tennessee 38104. *Received September 16, 1980.* This research was supported by the Veterans Administration and in part by U.S. Public Health Service Grants AA-03732 and AM-16506.

[1] Abbreviations used: CM-cellulose, carboxymethylcellulose; EDTA, ethylenediaminetetraacetic acid; PhNCS, phenylthiohydantoin; Staph V8 protease, a protease from *Staphylococcus aureus* VIII; Tris, 2-amino-2-(hydroxymethyl)-1,3-propanediol; TPCK, tosylphenylalanine chloromethyl ketone.

0006-2960/81/0420-2621$01.25/0

cording to established principles (Edman & Begg, 1967). The Slow Peptide-DMAA (071472) program of Beckman Instruments was 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. Subsequently, sequence analysis of overlapping peptides confirmed their identity.

Purification of Enzymatic and Hydroxylamine-Derived Peptides. Peptides obtained after cleavage of α1(III)-CB9 with trypsin, thermolysin, Staph V8 protease, and hydroxylamine were initially separated by gel filtration followed by ion-exchange chromatography and, in the case of tryptic peptides, 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. Four major molecular size fractions were collected. The smaller tryptic peptides (three to nine residues) were further separated by automated peptide analysis on a 0.9 × 20 cm column of PA-35 cation-exchange resin (Beckman Instruments, Palo Alto, CA) at 60 °C (Kang & Gross, 1970). An automatic peptide analyzer (Technicon Instruments, Inc., Ardsley, NY) equipped with a stream-splitting device allowed continuous monitoring of 20% of the effluent for reactivity with ninhydrin. The remaining 80% was collected in a fraction collector. Separation was achieved by utilizing a nine-chamber gradient device starting with 0.02 M sodium citrate buffer, pH 3.8 (Technicon peptide methodology). Ninhydrin-positive fractions were collected, their pHs were adjusted to 2.0 with 1 N HCl, and the fractions were desalted by using 1 × 2 cm columns of Aminex 50 × 8 (200–400 mesh) (Bio-Rad Laboratories, Richmond, CA) (Hirs, 1967). The two intermediate molecular size tryptic peptides obtained by gel filtration (11–30 residues) were separated by reverse-phase chromatography. Separation was obtained with a high-pressure liquid chromatograph (Waters Associates, Milford, MA) by the procedure of Rivier (1978) using an ODS-2 Magnum 9 reverse-phase column (Whatman, Inc., Bidwell, NJ). The initial solvent (A) consisted of 25% acetonitrile (Burdick & Jackson Laboratories) and 75% 0.01 M K_2HPO_4, pH 6.8; the limiting solvent (B) contained 75% acetonitrile and 25% 0.01 M K_2HPO_4, pH 6.8. 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% solvent B to 85% solvent B in 45 min. Peptides were continuously monitored by using a variable-wavelength spectrophotometer (Waters Instruments, Milford, MA) and collected in a fraction collector. Individual peptide fractions were collected and lyophilized, and a portion (10%) was taken for amino acid analysis. The remaining peptide (90%) was either adjusted to pH 4.0 with 1 M HCl and used for amino acid sequence analysis or, if necessary, desalted by Sephadex G-10 gel filtration in 0.1 M acetic acid and further purified by repeated reverse-phase chromatography using conditions identical with those described above. The largest tryptic peptide from Sephadex G-50S (mol wt ~7000) was the disulfide-linked trimer and was used without further purification.

The thermolytic reaction products of α1(III)-CB9 were separated by gel filtration with a Sephadex G-50S column in 0.1 M acetic acid. The two larger peptide fractions were further purified separately by a repeated Sephadex G-50S gel filtration in 0.1 M acetic acid. The largest of the thermolytic peptides was subsequently cleaved with trypsin, and the tryptic cleavage products were separated by Sephadex G-50S gel filtration to obtain the COOH-terminal tryptic peptide (mol wt ~7000) containing the cysteine residues (above). This peptide was reduced with 0.1 M dithiothreitol at 50 °C for 4 h in 0.1 M Tris at pH 7.4 with 0.001 M EDTA[1] and, subsequently, *S*-carboxymethylated with iodoacetic acid (Chung & Miller, 1974). The reduced, *S*-carboxymethylated peptide was again isolated by Sephadex G-50S gel filtration in 0.1 M acetic acid and used for amino acid sequence analysis.

The peptides obtained by cleavage with Staph V8 protease were fractionated initially on a Sephadex G-75S column (4.0 × 110 cm) equilibrated with 0.04 M sodium acetate, pH 4.8. The two major UV-absorbing peaks were separately subjected to CM-cellulose chromatography. A 2.5 × 20 cm column of CM-cellulose was equilibrated with 0.005 M sodium acetate, pH 4.8, and 43 °C, and after application of the sample a linear gradient of 0–0.3 M NaCl (total volume 1L) was used in each case to elute the peptides. The peptides isolated from CM-cellulose were further purified by Sephadex G-50S gel filtration in 0.1 M acetic acid prior to amino acid analysis and sequential degradation.

The peptide fragments of α1(III)-CB9 obtained after cleavage with hydroxylamine were separated by Sephadex G-75S gel filtration, and after reduction and *S*-carboxymethylation, the peptides were rechromatographed on Sephadex G-75S. Amino acid sequence analysis of only one reducible, *S*-carboxymethylated, HA-derived peptide was necessary to provide the needed information.

Results

Isolation of α1(III)-CB9. Intact type III collagen chains were prepared by CM-cellulose chromatography as previously described (Seyer et al., 1976, 1977). The disulfide-linked trimer of α1(III)-CB9 was isolated from CNBr digests initially by CM-cellulose chromatography at pH 3.83 (Seyer et al., 1977). Due to low recoveries of the peptide by this method, later procedures utilized an initial separation of the 66000 mol wt trimer from the other type III CNBr peptides by agarose A1.5m gel filtration. The next largest peptide of the CNBr digests was α1(III)-CB5 (mol wt 22000), thereby providing a more rapid method for isolating large quantities of α1-(III)-CB9. Confirmation of its identity and homogeneity was provided by reduction of the trimer with dithiothreitol (0.1 M in 1 M $CaCl_2$ and 0.5 M Tris, pH 7.4) at 50 °C for 4 h and isolation of the reduced 22000 mol wt α1(III)-CB9 on a similar column of agarose A1.5m equilibrated with 0.05 M Tris/1 M $CaCl_2$, pH 7.4, containing 0.001 M dithiothreitol (Seyer et al., 1980).

Isolation of Tryptic Peptides. α1(III)-CB9 after digestion with trypsin yielded a heterogeneous mixture of peptides which were partially fractionated into four molecular size fractions by Sephadex G-50S gel filtration.[2] Eleven small peptides from the low molecular weight fraction (T1–T11),[3] containing three to nine residues were isolated by PA-35 ion-exchange chromatography.[2] Their amino acid compositions are presented in Table I. The smaller, intermediate-size peptides obtained

[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, S for Staph V8 protease peptides, and HA for hydroxylamine peptides.

Table I: Amino Acid Composition of α1(III)-CB9 and Trypsin-Derived Peptides of Type III Collagen from Human Cirrhotic Liver[a]

amino acid	α1(III)-CB9	T1	T2	T3	T4	T5[b]	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18	T19	T20	total
4-Hyp[c]	29.2 (30)	1.1 (1)	1.1 (1)		1.0 (1)					0.9 (1)			1.8 (2)	1.7 (2)	1.2 (1)	0.8 (1)	1.7 (2)	2.6 (3)	4.7 (5)	4.6 (5)	5.7 (6)	31
Asp	(13)	1.0 (1)	0.2	1.0 (1)		0.2	1.1 (1)	1.0 (1)	0.2	0.2	0.8 (1)		3.1 (3)	1.3 (1)	0.2	0.2	1.0 (1)	0.2	1.4 (1)	1.2 (1)	1.4 (1)	13
Thr	2.3 (2)												0.2	0.2			0.8 (1)	1.2 (1)	0.4	0.3	0.2	2
Ser	(11)	0.8 (1)					1.1 (1)						0.8 (1)	2.0 (2)	2.1 (2)	0.8 (1)	0.8 (1)	0.2	0.9 (1)	2.2 (2)	2.2 (2)	14
Glu	(15)		1.1 (1)		2.2 (2)		0.2		1.2 (1)	0.2			0.2	4.4 (4)	1.4 (1)	0.3	0.3	2.4 (2)	1.2 (1)	2.3 (2)	1.3 (1)	15
Pro[c]	26.1 (26)	0.2	2.3 (2)			0.2	0.2			1.2 (1)			0.4	2.3 (2)	3.4 (3)	3.3 (3)	3.4 (3)	2.3 (2)	3.4 (3)	3.4 (3)	3.4 (3)	25
Gly	(80)	3.4 (3)	3.0 (3)	1.1 (1)	2.1 (2)	2.0 (2)	1.2 (1)	1.2 (1)	1.0 (1)	1.0 (1)	1.0 (1)	1.0 (1)	4.1 (4)	8.2 (8)	6.1 (6)	4.2 (4)	5.1 (5)	6.9 (7)	8.6 (9)	9.0 (9)	9.8 (10)	80
Ala	(19)		1.2 (1)		0.2	2.1 (2)	0.3	0.2					0.2	1.0 (1)	4.3 (4)	0.3	1.0 (1)	2.3 (2)	3.2 (3)	4.1 (4)	1.0 (1)	19
Cys[d]	1.5 (2)																				1.6 (2)	2
Val	3.8 (4)													0.7 (1)		0.6 (1)		0.2	0.6 (1)	0.2	0.6 (1)	4
Ile	2.7 (3)					0.7 (1)											0.7 (1)	0.3		0.9 (1)	0.2	3
Leu	3.3 (3)												1.1 (1)		0.2		0.2	1.8 (2)	0.2		0.3	3
Phe	1.2 (1)																	0.2	0.2	1.0 (1)		1
Hyl[c]	2.8 (3)					0.6 (1)								0.7 (1)					0.2			2
Lys[c]	4.3 (4)	0.8 (1)		0.8 (1)	0.2	0.2								1.1 (1)	0.2	0.8 (1)	0.2	0.3	1.1 (1)			5
His	2.8 (3)											1.1 (1)							0.7 (1)	0.2	0.8 (1)	3
Arg	(14)		1.0 (1)	0.2	1.0 (1)		1.1 (1)	0.9 (1)	0.8 (1)	1.0 (1)	1.1 (1)	0.9 (1)	1.0 (1)	1.1 (1)	0.9 (1)	0.2	0.9 (1)	1.0 (1)	0.2	0.9 (1)	0.2	14
total	234	7	9	3	6	6	4	3	3	4	3	3	12	24	18	11	16	20	26	30	28	236

[a] Values are expressed as residues per peptide. No entry indicates the level was <0.1 residue/peptide. Numbers in parentheses indicate assumed integral values. [b] Two peptides (Gly-Asx-Arg) were isolated. T7 contained the sequence Gly-Asp-Arg, while T10 contained the sequence Gly-Asn-Arg. [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 the amino acid content of α1(III)-CB9 when analyzed by amino acid analysis and by sequence studies. [d] Quantitated as *S*-(carboxymethyl)cysteine.

by Sephadex G-50S gel filtration[2] contained tryptic peptides of 11–24 residues including some uncleaved peptides. These were separated by reverse-phase high-pressure liquid chromatography[2] into six peptides (T12–T17). Their amino acid compositions are listed in Table I. The larger intermediate molecular weight fraction isolated by Sephadex G-50S gel filtration was separated into two tryptic peptides (26 and 30 residues) by reverse-phase chromatography.[2] The amino acid compositions of these (T18, T19) are also tabulated in Table I. The largest peptide (T20)[2] obtained from Sephadex G-50S gel filtration eluted as a disulfide-linked tryptic peptide, $(T20)_3$, mol wt ~7000, which on subsequent reduction, *S*-carboxymethylation, and a repeated gel filtration, eluted as the monomeric peptide. The sum of the amino acid compositons of the above 20 tryptic peptides accounts for the entire amino acid content of intact α1(III)-CB9 within experimental error. The tryptic peptide T20 contained no basic residue or homoserine and hence was established as the COOH terminus of the helical (pepsin-resistant) type III collagen. Two residues of *S*-(carboxymethyl)cysteine were also found in this peptide.

Since there are 21 basic residues in α1(III)-CB9, one might have anticipated 22 tryptic peptides. However, one of the peptides, T13, contained a residue each of lysine, hydroxylysine, and arginine. The reason for tryptic cleavage of Hyl in T5 but not T13 is unclear, but previous studies have suggested varying susceptibility of Hyl residues to tryptic digestions (Seyer & Kang, 1978). The uncleaved lysine (T13) may be explained by the fact that it was followed by Hyp (see below). At pH 8.0, all Lys-Hyp bonds have previously been found resistant to tryptic cleavage. Thus, in summary, the 20 tryptic peptides obtained account for the entire amino acid content of α1(III)-CB9 (Table I).

Alignment of Tryptic Peptides. The alignment of each tryptic peptide of α1(III)-CB9 was deduced from automated Edman degradation of the intact peptide, each of the tryptic peptides, thermolytic peptides, Staph V8 protease peptides, and one of the hydroxylamine peptides. Sequential degradation of intact α1(III)-CB9 through the first 39 residues from the NH_2 terminus provided the necessary information to align the tryptic peptides, T9–T13–T17, and locate the uncleaved Hyl (residue 13) and Lys–Hyp sequence (residue 18–19) in T13 (Figure 1). Additional information was obtained by characterization of thermolytic peptides of α1(III)-CB9. The thermolytic digest of α1(III)-CB9 was initially separated by gel filtration on Sephadex G-50S.[2] Two major high molecular weight (>14 000 and 10 000) fractions (Th1, Th2) were obtained. Each fraction was separately further purified by a second gel filtration using identical conditions (not shown). Although other smaller molecular weight peptides were present, only these, Th1 and Th2, were necessary to establish the alignment of the tryptic peptides. Th1 represented the COOH terminus of α1(III)-CB9 as evidenced by its amino acid composition (Table II), including two residues of cysteine/mol, and isolation of T20 from a tryptic digest of Th1 (see below). The thermolytic peptide Th2 contained Leu as its NH_2 terminus, and subsequent automated Edman degradation of Th2 and T17 established that Th2 contained the COOH-terminal portion of T17, with thermolytic cleavage at the Gly–Leu bond (residue 38–39). Subsequent analysis of the initial 42 residues of Th2 provided the necessary information to align tryptic peptides T17–T12–T1–T7–T18 (Figure 1) and confirmed the previous sequence analysis of each of the tryptic peptides. The amino acid composition of Th2, presented in Table II, was not useful in identifying the exact location of the COOH terminus of Th2, but T5 appeared to be the sen-

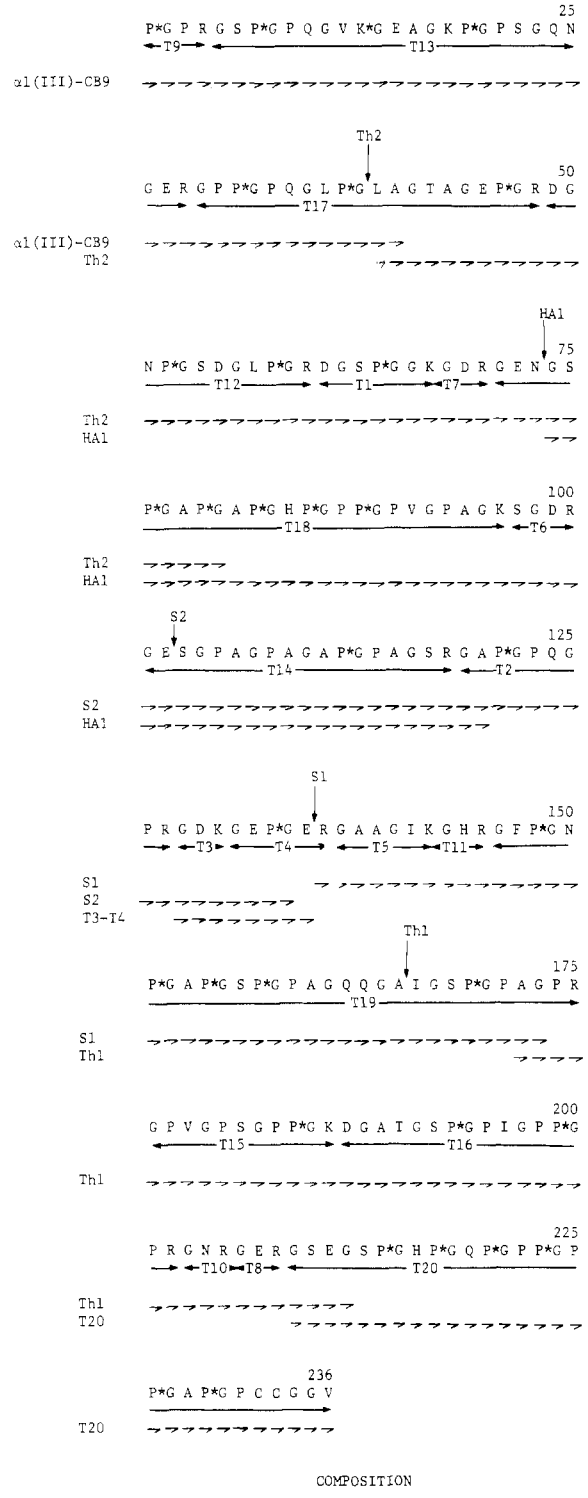

FIGURE 1: Complete amino acid sequence of α1(III)-CB9 of human liver type III collagen. The tryptic peptides were isolated as indicated by long arrows (↔). Short arrows (→) indicate the extent of Edman degradation of each peptide. All tryptic peptides were analyzed although only the uncleaved T3–T4 and T20 were presented in this figure to indicate the complete sequence analysis. The symbols S1, S2, Th1, Th2, and HA1 denote the cleavage sites of Staph V8 protease, themolysin, and hydroxylamine, respectively.

sitive peptide since no Ile (residue 141) was present. Its amino acid composition would be consistent with cleavage at this Gly–Ile bond. The thermolytic peptide Th1 contained the two residues of cysteine and was identified as the COOH-terminal portion of α1(III)-CB9. Automated Edman degradation of Th1 revealed Ile (residue 166) as the NH_2 terminus followed by the sequence of the COOH-terminal portion of 19. Forty-seven amino acid residues were identified from the NH_2 terminus of Th1 which produced data sufficient to align the COOH-terminal tryptic peptides T19–T15–T16–T10–T8–T20.

Enzymatic digestion with Staph V8 protease was used to selectively cleave α1(III)-CB9 at the COOH side of specific Glu residues. The digestion mixture was separated by Sephadex G-75S gel filtration.[2] Two major molecular size fractions were obtained with molecular weights of approximately 8000 and 4000, respectively, and each was further purified by CM-cellulose chromatography at pH 4.8. The larger fraction (8000 mol wt) contained peptides which eluted with a 0–0.2 M NaCl linear gradient and a 0.5 M NaCl wash. The peptide S1 was subjected to automated Edman degradation, and a total of 37 residues were identified (Figure 1). From this data, the alignment of tryptic peptides Arg–T5–T11–T19 which contained the thermolytic cleavage site of Th1 was determined. The amino acid composition of S1, if cleavage at a Glu–Arg bond (residue 135–136) of T4 is assumed, suggested a 72-residue peptide with possible cleavage (COOH terminus) at Glu–Arg (residue 206–207) in T16. Another Staph V8 protease peptide useful in determining the alignment of the tryptic peptide S2 was obtained by CM-cellulose chromatography[2] of the smaller molecular weight fraction from Sephadex G-75S chromatography. Sequential degradation of S2 showed that Ser was the NH_2 terminus followed with 32 amino acid residues having a sequence consistent with the alignment of T14–T2–T3–T4 (Figure 1). The COOH terminus of S2 was tentatively located in the tryptic peptide T4 (Gly-Glu-Hyp-Gly-Glu-Arg). The amino acid composition of S2 is presented in Table I and supports this conclusion. Alignment of the T3–T4 was confirmed by the isolation of the uncleaved tryptic peptide, Gly-Asp-Hyl-Gly-Glu-Hyp-Gly-Glu-Arg obtained from reverse-phase chromatography.[2] The Hyl residue was apparently only partially hydroxylated since a Gly-Asp-Lys (T3) was also isolated by automatic peptide ion-exchange separation[2] (above).

The remaining alignment was deduced from analyses of hydroxylamine cleavage products. The peptide fragments were initially separated by Sephadex G-75S gel filtration, and a major component was eluted as an excluded peak (>24 000). Reduction and *S*-carboxymethylation of this fraction and gel filtration again using identical methods yielded HA1 (not shown) which now eluted at a position corresponding to a 15 000 mol wt peptide. Further purification by CM-cellulose chromatography and automated Edman degradation identified the amino terminal sequence of Gly-Ser-Hyp-Gly-Ala-Hyp (Figure 1). Identification of 47 residues of HA1 provided information for the T18–T6–T14 alignment. The amino acid composition of HA1 also revealed the presence of two *S*-(carboxymethyl)cysteine residues/mol and must, therefore, represent the COOH-terminal hydroxylamine-derived fragment of α1(III)-CB9.

In summary, the peptide alignment obtained by a combination of automated Edman degradation of intact α1(III)-CB9, thermolytic peptides Th1 and Th2, Staph V8 protease peptides S1 and S2, and the hydroxylamine fragment HA1 provided the tryptic peptide sequence of T9–T13–T17–T12–T1–T7–T18–T6–T14–T2–T3–T4–T5–T11–T19–T15–T16–T10–T8–T20. The use of specific peptides Th1, Th2, S1, S2, and HA1 made this peptide alignment possible as demon-

Table II: Amino Acid Composition of Thermolytic, Staph V8 Protease, and Hydroxylamine Peptides Derived from α1(III)-CB9 of Human Cirrhotic Liver Collagen[a]

amino acid	position:[b] peptide:	165–212 Th1	39–136 Th2	135–203 S1	103–133 S2	74–236 HA1
4-Hyl[c]		8.1 (8)	10.6 (11)	8.1 (8)	2.6 (3)	22.1 (23)
Asp		2.3 (2)	8.3 (8)	3.3 (3)	1.3 (1)	5.8 (6)
Thr		0.8 (1)	0.9 (1)	0.9 (1)	0.2	0.6 (1)
Ser		4.6 (5)	5.8 (6)	4.1 (4)	0.9 (1)	9.6 (10)
Glu		3.4 (3)	6.3 (6)	3.0 (3)	3.4 (3)	8.3 (8)
Pro		13.2 (13)	8.0 (8)	8.1 (8)	4.6 (5)	19.4 (19)
Gly		24.0 (24)	32.0 (32)	23.0 (23)	10.0 (10)	55.4 (56)
Ala		3.0 (3)	11.0 (11)	7.2 (7)	5.0 (5)	15.8 (16)
Cys/2[d]		1.6 (2)				1.6 (2)
Val		1.6 (2)	0.7 (1)	0.9 (1)		2.6 (3)
Ile		1.9 (2)	0.2	2.6 (3)	0.2	3.1 (3)
Leu		0.3	2.2 (2)	0.3	0.3	0.2
Phe				1.1 (1)		0.8 (1)
Hyl		0.3	0.6 (1)	0.8 (1)	0.6 (1)	0.8 (1)
Lys		0.9 (1)	2.3 (2)	1.4 (1)	0.4	3.2 (3)
His		0.8 (1)	0.6 (1)	0.8 (1)		2.6 (3)
Arg		4.3 (4)	7.2 (7)	3.1 (3)	1.0 (1)	9.3 (9)
total		71	97	68	30	163

[a] Values are expressed as residues per peptide. No entry indicates the level was <0.1 residue/peptide. Integral values were given for residues present in >10 residues. Numbers in parentheses indicate assumed integral values. [b] Location of the peptide in α1(III)-CB9. [c] Partial proline and lysine hydroxylation was noted by sequence analysis. Actual values were therefore given for these amino acids. [d] Determined as *S*-(carboxymethyl)cysteine.

strated in Figure 1. Final identification of the Cys-Cys sequence of α1(III)-CB9 was made after *S*-carboxymethylation of the tryptic peptide T20.

Internal Sequence of Various Peptides. All peptide were modified at the COOH terminus by treatment with 2-amino-1,5-naphthalenedisulfonic acid in the presence of *N*-ethyl-*N'*-[3-(dimethylamino)propyl]carbodiimide. Each peptide was degraded at least twice. The complete sequence of α1(III)-CB9 obtained in the present study is summarized in Figure 1. Larger peptides were used to obtain the necessary overlap of the tryptic peptides. Additional data including yields of the automated Edman degradation of peptides are summarized in supplementary material.[2] Only the minimal data required for deduction of the complete sequence of α1-(III)-CB9 were included in this paper.

Discussion

Type III collagen was obtained by limited pepsin digestion of human cirrhotic liver. The covalent structure of α1-(III)-CB9 was first initiated by isolation and primary structural studies with the 20 tryptic peptides. Two thermolysin-sensitive bonds (Gly–Leu and Ala–Ile) at positions 38–39 and 165–166, two Staph V8 protease sensitive bonds (Glu–Ser and Glu–Arg) at positions 102–103 and 135–136, and the hydroxylamine cleavage of Asn–Gly bond (residue 73–74) enabled completion of the tryptic peptide alignment and complete amino acid sequence of the 236 residues. The occurrence of a Cys-Cys peptide sequence in type III collagen represents the termination of the helical, Gly-X-Y type structure with the second Cys residue replacing the usual Gly at that location. The pepsin-sensitive cleavage area of the type III molecule was apparently after Val in the Cys-Cys-Gly-Gly-Val sequence of T20.

The remaining sequence of α1(III)-CB9 is similar in most respects to other collagen primary structures thus far identified (Hulmes et al., 1973; Gallop & Paz, 1975; Fietzek & Kuhn, 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 contrast to other type III collagen peptides thus far identified which contained a considerable excess of Gly in either the X or Y position, only one excess Gly residue was found and identified as a Gly-Gly-Lys sequence of T1. Twenty-two excess Gly residues were found in the previous 795 amino acid segment. Pro in the X position is only hydroxylated at the single Gly-3-Hyp-4-Hyp sequence of the COOH region of the

Table III: Distribution of Amino Acids in Gly-X-Y Positions of Human Liver Type III Collagen and Frequency of Substitutions with Human Liver and Calf Skin Type III Collagen

amino acid		X	Y	total	no. of substitutions[a]	% of substitutions[b]
Hyp	(P*)	0	146[c]	146	12	4.1
Asp	(D)	9	17	26	6	11.6
Asn	(N)	8	14	22	6	13.6
Thr	(T)	6	8	14	8	28.6
Ser	(S)	29	10	39	29	37.2
Glu	(E)	46	2	48	4	4.2
Gln	(Q)	10	17	27	5	9.2
Pro	(P)	96	0	96	12	6.2
Gly	(G)[d]	15	4	360	12	31.6
Ala	(A)	46	44	90	42	23.3
Cys/2	(C)	0	1	1	0	0.0
Val	(V)	8	3	11	13	59.1
Met	(M)	5	4	9	6	33.3
Ile	(I)	10	4	14	12	42.9
Leu	(L)	18	1	19	12	31.6
Tyr	(Y)	2	0	2	0	0.0
Phe	(F)	8	0	8	1	6.3
Lys	(K)	11	26[c]	37	1	1.4
His	(H)	6	0	6	2	16.7
Arg	(R)	8	40	48	1	1.0
total		341	341	1023	184	

[a] Both human liver and calf skin type III collagen chains were used as parent molecules such that Hyp, for instance, in either species, when substituted, was tabulated. The total number presented was therefore 2-fold greater than that expected if one collagen chain was used as the parent collagen. [b] The values were obtained by utilizing the total of each specific amino acid in human liver type III collagen and the number of substitutions of combined human liver and calf skin type III collagen. Since both chains were used in tabulating the number of substitutions, the percentage obtained was reduced by half to obtain percent of substitution. [c] The partial hydroxylation was considered fully hydroxylated for this computation. [d] A total of 341 triplets were present; hence, there were 341 + 15 + 4 = 360 glycine residues/1023 residues. Substitution was computed only with Gly residues in the X and Y position since the initial 341 Gly residues were invariant.

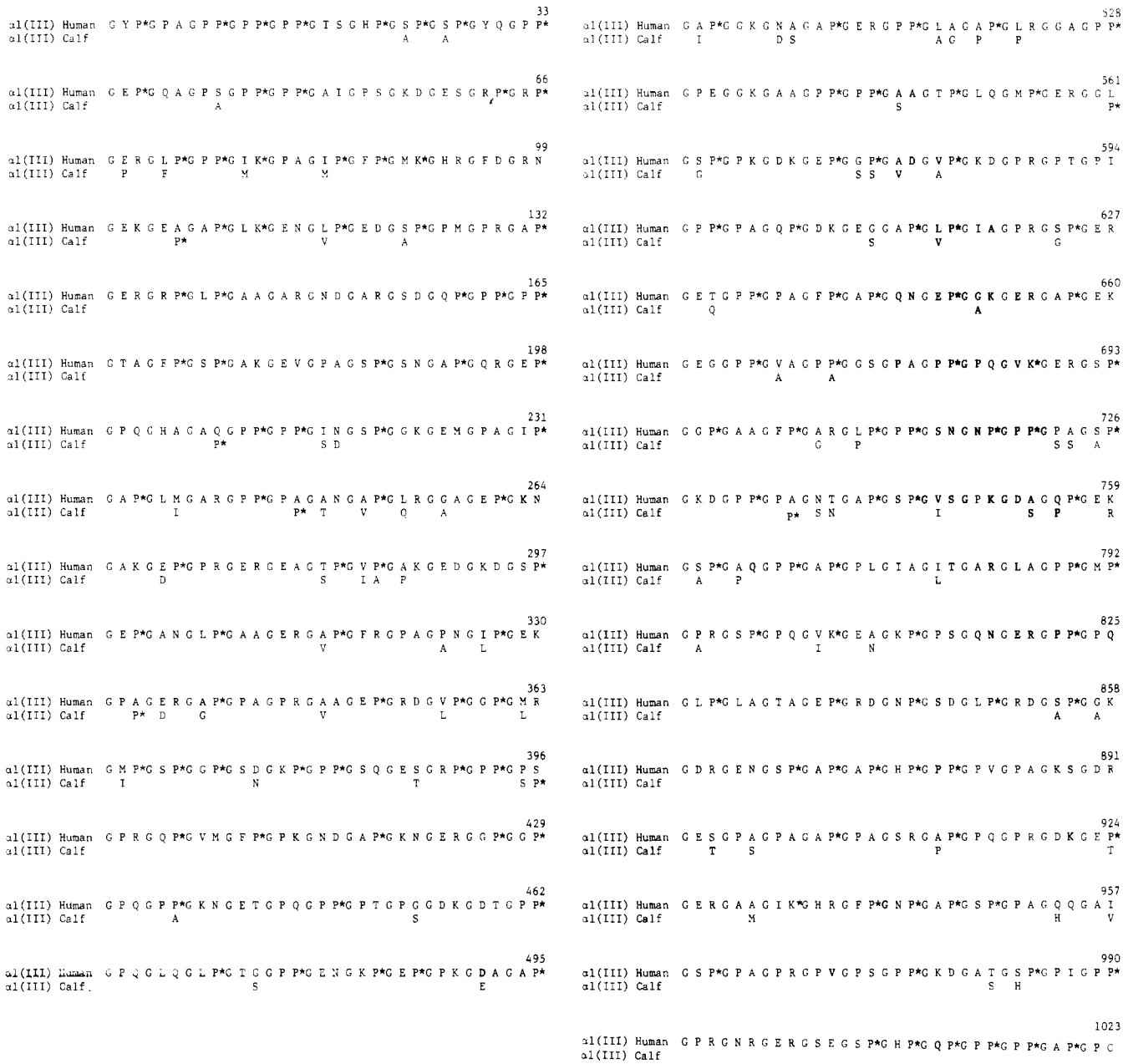

FIGURE 2: Comparison of amino acid sequences of helical portion of human cirrhotic liver type III collagen with calf skin type III collagen. Identical residues in the same position of each chain are indicated by a blank space. The numbering system was initiated with the first helical Gly-X-Y triplet and ended with the last triplet containing a Cys residue in the Y position. Nonhelical regions of the NH_2- and COOH-terminal segments of the molecule including the additional Gly-Ile-Ala triplet preceding the NH_2-terminal Gly-Tyr-Hyp of calf skin type III collagen chain were not included.

α1(I) collagen chain (residue 1002), but no 3-Hyp was found in the present peptide even though Pro was present in the homologous site (residue 198, Figure 1), thereby making the role of 3-Hyp in type I collagen uncertain. Prolines in the Y positions were generally >90% hydroxylated except for the residue 184 which was estimated to be 50% hydroxylated. The Lys residue, as stated above, contained varying degrees of hydroxylation when present in the Y position. The Lys residues at 13 and 187 were fully hydroxylated, and Lys at 180 was estimated to be 50% hydroxylated.

The complete structure of the helical region of human cirrhotic liver type III collagen together with the primary structure of bovine skin type III collagen is presented (Figure 2). The numerical system used was initiated with the pepsin-resistant, helical Gly-Tyr-Hyp sequence since the preceding three amino acids were Val-Leu-Ala (Seyer & Kang, 1977). The recently identified calf skin type III collagen contains an additional Gly-X-Y triplet (Gly-Ile-Ala) preceding the Gly-Tyr-Hyp sequence and therefore contains one triplet more than the human liver type III collagen. Except for a single Lys → Arg substitution, complete homology was found in every Lys and Arg residue. Phe was substituted only once with Leu being present in human type III collagen. Tyr was located at two positions near the NH_2 terminus. The bovine type III collagen contained two additional His residues, and as stated previously, no 3-Hyp was found in either species. Ninety-two amino acid substitutions of the 1023 amino acids were identified, giving 91% homology. The hydrophobic amino acids Val, Ile, and Leu were highly substituted, but this was generally to another hydrophobic residue. Pro and Hyp represent 242 amino acids, yet substitutions were relatively rare (4.1 and 6.2%, respectively) compared with the 90 Ala residues (23% substitutions). The acidic amino acid substitutions Glu and Asp were also limited (4.2 and 11.6%, respectively). Table III contains the

content of each amino acid per chain, its location in the Gly-X-Y triplet, and the frequency of substitution. Consistent with similar comparisons of α1(I) collagen chains, the basic residues Arg and Lys occur more frequently in the Y position while hydropholic residues prefer the X position. Tyr and His were found only in the X position, and Glu was located primarily in the X position while Gln was present mainly in the Y position.

In summary, the complete primary structure of the helical region of human liver type III collagen has been completed. The present report encompasses the COOH-terminal 236 residues including the additional Cys-Gly-Gly-Val sequence which was apparently resistant to the limited pepsin digestion used to initially solubilize human type III collagen. The Hyl-Gly-His-Arg tetrapeptide sequence was repeated twice in homologous positions as in type I collagen, which would allow an identical quarter stagger structure and lysine or hydroxylysine aldehyde derived cross-linkages between NH_2 or COOH nonhelical regions with Hyl or Lys in the helical Hyl-Gly-His-Arg segment of either α1(III)-CB7 or α1-(III)-CB9. The significance of the Cys-Cys sequence at the COOH terminus and their subsequent stabilizing effect with disulfide intramolecular cross-linkages remains uncertain. As reported previously, collagenase cleavage of type III collagen occurs one triplet prior to the specific collagenase cleavage site of type I collagen (Seyer et al., 1980). Basically, the type III collagen of human liver was homologous in most respects thus far examined to type I collagen, and this leads to the conclusion that they are genetically related although the biological significance of two separate collagens in many interstitial tissues remains to be determined.

Acknowledgments

We 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, M. Shenk, and T. Todd. Sincere thanks are given to A. Bobrowski and H. Oellerich for typing the manuscript.

Supplementary Material Available

One table showing yields of automated Edman degradation of peptide fragments of α1(III)-CB9 from human liver type III collagen and eight figures showing isolation of peptides of α1(III)-CB9 by gel filtration, ion-exchange chromatography, reverse-phase chromatography, and CM-cellulose chromatography (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.

Chung, E., & Miller, E. J. (1974) *Science (Washington, D.C.) 183*, 1200.

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. O., Corbett, C., Kang, A. H., & Gross, J. (1975) *Biochemistry 14*, 1933.

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

Fietzek, P. P., & Kühn, K. (1976) *Int. Rev. Connect. Tissue Res. 7*, 1.

Fietzek, P. P., Allman, 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.

Hirs, C. H. W. (1967) *Methods Enzymol. 11*, 386.

Hulmes, D. J. S., Miller, A., Parry, D. A., Piez, K. A., & Woodheat-Galloway, J. (1973) *J. Mol. Biol. 70*, 137.

Kang, A. H. (1972) *Biochemistry 11*, 1828.

Kang, A. H., & Gross, J. (1970) *Biochemistry 9*, 796.

Lang, H., Glanville, R. W., Fietzek, P. P., & Kühn, K. (1979) *Hoppe-Seyler's Z. Physiol. Chem. 360*, 841.

Piez, K. A. (1976) *Biochemistry* of Collagen (Ramachandran, G. N., & Reddi, A. H., Eds.) Plenum Press, New York.

Rivier, J. E. (1978) *J. Liq. Chromatogr. 1*, 343.

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.

Seyer, J. M., Mainardi, C. L., & Kang, A. H. (1980) *Biochemistry 19*, 1583.

Smithies, O., Gibson, D., Fanning, E. M., Goodfliesh, R. M., Gilman, J. G., & Ballantyne, D. L. (1971) *Biochemistry 10*, 4912.

Zimmerman, C. L., Pisano, J. J., & Appella, E. (1973) *Biochem. Biophys. Res. Commun. 55*, 1220.